

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:13 ; Search time 42 Seconds

(without alignments)

1900.408 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASPTATAAVSGRWLGNGHT.....VPYGFHGFICADSLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3150	100.0	599	AAE04784	Arabidopsis thaliana
2	3150	100.0	599	ABE92311	Herbicidally activ
3	2280.5	72.4	605	AAE04789	Lycopersicon escul
4	2280.5	72.4	605	AAE04787	Neoxanthin cleavag
5	2168.5	68.8	612	AAE04787	Vigna unguiculata
6	2044.5	64.9	589	ABE91182	Herbicidally activ
7	1991	63.2	583	AAE04782	Arabidopsis thaliana
8	1930	61.3	604	AAE04788	Zea mays neoxanthi
9	1930	61.3	604	AAE04788	Neoxanthin cleavag
10	1663.5	52.8	577	AAE04786	Arabidopsis thaliana

11	1663.5	52.8	577	23	ABB92416	Herbicidally activ
12	968	30.7	595	21	AAE04784	Arabidopsis thaliana
13	968	30.7	595	22	AAE04783	Arabidopsis thaliana
14	968	30.7	595	22	AAE04783	Neoxanthin cleavag
15	959	30.4	517	21	AAE04783	Arabidopsis thaliana
16	939	29.8	538	21	AAE04783	Arabidopsis thaliana
17	939	29.8	539	21	AAE04783	Arabidopsis thaliana
18	938	29.8	538	22	AAE04785	Arabidopsis thaliana
19	937	29.7	538	22	AAE04790	Arabidopsis thaliana
20	925.5	29.4	501	21	AAE04790	Arabidopsis thaliana
21	857.5	27.2	544	22	AAE04790	Sunflower neoxanth
22	834.5	26.5	431	21	AAE04790	Arabidopsis thaliana
23	745	23.7	446	22	AAE04790	Neoxanthin cleavag
24	322.5	10.2	570	22	AAE04790	Arabidopsis thaliana
25	307.5	9.8	596	22	AAE04790	Arabidopsis thaliana
26	279	8.9	177	22	AAE04790	Physcomitrella pat
27	272.5	8.7	166	21	AAE04790	Zea mays protein f
28	264.5	8.4	153	21	AAE04790	Zea mays protein f
29	260	8.3	526	21	AAE04790	Beta, Beta-caroten
30	259.5	8.2	152	21	AAE04790	Zea mays protein f
31	256	8.1	516	22	AAE04790	Amino acid sequenc
32	247	7.8	506	21	AAE04790	Beta, Beta-caroten
33	238.5	7.6	549	22	AAE04790	Zebra-2, beta-car
34	232.5	7.4	532	22	AAE04790	Mouse-2, beta-car
35	223	7.1	547	22	AAE04790	Human protein sequ
36	210	6.7	533	22	AAE04790	Canine RPE65, Can
37	209.5	6.7	539	22	AAE04790	Human protein sequ
38	209	6.6	533	14	AAE04790	Human retinol bind
39	207.5	6.6	556	22	AAE04790	Human-2, beta-car
40	206.5	6.6	579	22	AAE04790	Human RECAC poly
41	185.5	5.9	529	21	AAE04790	Beta, Beta-caroten
42	154.5	4.9	65	23	AAE04790	Human ORFX protein
43	150	4.8	620	22	AAE04790	Drosophila melanog
44	150	4.8	620	22	AAE04790	Drosophila beta-ca
45	150	4.8	620	22	AAE04790	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAE04784
ID AAE04784 standard; Protein; 599 AA.

AC AAE04784;
XX
XX
DT 10-SEP-2001 (first entry)
XX

XX Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCED3.

DE Neoxanthin cleavage enzyme; ATNCED3; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant.

XX Arabidopsis thaliana.

XX Ep1116794-A2.

PD 18-JUL-2001.

PF 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI: 2001-400081/43.

XX N-PSDB; AAD09396.

XX A DNA encoding a protein with a neoxanthin cleavage activity for

PT producing transgenic plants with improved or decreased stress tolerance
 PT
 XX
 PS
 PS
 XX
 XX
 CC Claim 3; Fig 9; 101pp; English.
 CC
 CC The invention relates to neoxanthin cleavage enzymes and their
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
 CC plant when expressed in a plant cell. The invention also relates to
 CC methods for increasing or decreasing stress tolerance in a plant by
 CC introducing the DNA into the plant, and a transgenic plant into which a
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid
 CC land can be improved by growing transformant weed for several years and
 CC then removing the weed by specifically lowering stress tolerance in the
 CC weed by inducing an inducible promoter. The present sequence is
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
 CC The AtNCE3 cDNA is obtained from an Arabidopsis plant-derived cDNA
 CC library using a cDNA of the cPRD65 (CovPea Responsive to Dehydration)
 CC gene isolated from cowpea plant as a probe.

XX
 SQ Sequence 599 AA;
 Query Match 100.0%; Score 3150; DB 22; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.6e-310;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASFTATAAAGSGRWLGNGHTOPPLSSOSSDLSYCSSLPMAASRVTRKLNVSALHTPPAL 60
 DB 1 MASFTATAAAGSGRWLGNGHTOPPLSSOSSDLSYCSSLPMAASRVTRKLNVSALHTPPAL 60
 QY 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 DB 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 QY 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 QY 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGIVDP 240
 QY 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTGVRDFDGLSTMTIAHPKV 300
 DB 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTGVRDFDGLSTMTIAHPKV 300
 QY 301 DPESGELFALSVDVSKPKYKFRSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 DB 301 DPESGELFALSVDVSKPKYKFRSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 QY 361 QVVFKLPEMIRGSPVVDKKNVAREGILDYKVAEDSSNIKWIDAPDCFCFHLNAAWEPE 420
 DB 361 QVVFKLPEMIRGSPVVDKKNVAREGILDYKVAEDSSNIKWIDAPDCFCFHLNAAWEPE 420
 QY 421 TDEVVVGSCMTPPDSIFNSEDENLKSVLSEIRLNKLTGESIRRPITISNEDQOVNLEAGM 480
 DB 421 TDEVVVGSCMTPPDSIFNSEDENLKSVLSEIRLNKLTGESIRRPITISNEDQOVNLEAGM 480
 QY 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGE 540
 QY 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSPVPGFHCIFICADDLAKOVV 599
 DB 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEATVKLPSPVPGFHCIFICADDLAKOVV 599

RESULT 2
 ABB92311
 ID ABB92311 standard; Protein; 599 AA.

XX
 AC ABB92311;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 1522.
 XX
 KW Herbicidal; plant; agriculture; herbicide.
 XX
 XX Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 XX 28-AUG-2001; 2001WO-EP09892.
 XX
 XX 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (PARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 XX
 DR WPI; 2002-269010/31.
 XX
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 1522; 26lpp + Sequence Listing; English.
 CC
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 XX
 SQ Sequence 599 AA;
 Query Match 100.0%; Score 3150; DB 23; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.6e-310;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASFTATAAAGSGRWLGNGHTOPPLSSOSSDLSYCSSLPMAASRVTRKLNVSALHTPPAL 60
 DB 1 MASFTATAAAGSGRWLGNGHTOPPLSSOSSDLSYCSSLPMAASRVTRKLNVSALHTPPAL 60
 QY 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 DB 61 HFPKSSNSPAIVVVKPKAKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLKPTADPS 120
 QY 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VOIAGNFAPVNEQPVRRNLPPVVGKLPDSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 QY 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASYACRFTQTNRFVQERQLGRPVFPAIGELHGHGTGIARLMLFYARAAAGIVDP 240
 QY 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTGVRDFDGLSTMTIAHPKV 300
 DB 241 AHGTGVANAGLYVFNGRLLAMSEDDLPYQVQITPNGDLKTGVRDFDGLSTMTIAHPKV 300
 QY 301 DPESGELFALSVDVSKPKYKFRSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 DB 301 DPESGELFALSVDVSKPKYKFRSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360
 QY 361 QVVFKLPEMIRGSPVVDKKNVAREGILDYKVAEDSSNIKWIDAPDCFCFHLNAAWEPE 420

```

|||||
Db 361 QVVKLPKPEMLRGSPVYDANKVARFGLIDKYAEDSSNIKWIDAPDFCFCHLNWANEPE 420
|||||
Qy 421 TDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIISNEDQQVNLBAGM 480
|||||
Db 421 TDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIISNEDQQVNLBAGM 480
|||||
Qy 481 VNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPFLPUGGGGE 540
|||||
Db 481 VNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPFLPUGGGGE 540
|||||
Qy 541 EDEGYIILCFVHDEKTKWSELOQVNAVSELEVEATVKLSRPVYGHGTFIGADDLAKQVY 599
|||||
Db 541 EDEGYIILCFVHDEKTKWSELOQVNAVSELEVEATVKLSRPVYGHGTFIGADDLAKQVY 599
|||||

RESULT 3
AAE04789
ID AAE04789 standard; Protein; 605 AA.
XX
AC AAE04789;
XX
DT 10-SEP-2001 (first entry)
XX
Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCEd1.
XX
Tomato; neoxanthin cleavage enzyme; LeNCEd1; abscisic acid; ABA;
XX
stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX
plant growth protectant; herbicide.
XX
Lycopersicon esculentum.
XX
EP116794-A2.
PN
XX
18-JUL-2001.
PD
XX
11-JAN-2001; 2001EP-0300218.
PF
XX
13-JAN-2000; 2000JP-0010056.
PR
11-JAN-2001; 2001JP-0003476.
XX
XX
(RIKE ) RIKEN KK.
XX
Iuchi S, Kobayashi M, Shinozaki K.
XX
WPI: 2001-400081/43.
DR
N-PSDB; AAD09401.
XX
A DNA encoding a protein with a neoxanthin cleavage activity for
producing transgenic plants with improved or decreased stress tolerance
XX
Claim 3; Fig 2; 101pp; English.
XX
The invention relates to neoxanthin cleavage enzymes and their
corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
Neoxanthin cleavage enzyme is used for improving stress tolerance in a
plant when expressed in a plant cell. The invention also relates to
methods for increasing or decreasing stress tolerance in a plant by
introducing the DNA into the plant, and a transgenic plant into which a
neoxanthin cleavage enzyme is introduced. The improvement of stress
tolerance in plants is useful, for example in plant breeding. Neoxanthin
cleavage enzyme genes are useful for producing transgenic plants. An arid
land can be improved by growing transformant weed for several years and
then removing the weed by specifically lowering stress tolerance in the
weed by inducing an inducible promoter. The present sequence is
CC Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCEd1 protein
CC related to the invention.
XX
XX
Sequence 605 AA;
XX
Query Match 72.4%; Score 2280.5; DB 22; Length 605;

```

```

Best Local Similarity 71.7%; Pred. No. 53e-222;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

Qy 5 TATAAAGCGMLGGNHTQPIPLSSOSSDLSYCS---SLPMASRVTRKLNVSALHTPALH 61
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 TTTSHATNTWT-----KTKLSMPSSKEFFGAFASISLILKNQHRQSLNINSLSQAPLII 57
|||||
Qy 62 FPKOSSN--SPA--IVVKPKAKESN-----TKQNNLFQRAAAALDRAEGFLVSHKELHP 112
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 FPKOSSNYQTPKNTTISHPKOENNNSSSSTSKNWLQKAAAMALDAVESALTKEHEHP 117
|||||
Qy 113 LPTADPSVQIAGNAPVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHPBPVTHGHFFD 172
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 LPTADPRVQISGNAPVNPVPCOSLPVTKIPKCVQGVYVRNGANPLFEPTAGHFFD 177
|||||
Qy 173 GDGMVHAVKFEHGSASACRYTQNRVQERQLGRPVFPKAIKELHGHGTIARLMLFYAR 232
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 GDGMVHAVQFKNGSASACRYTETERLVQEKALGRPVFPKAIKELHGHGTIARLMLFYAR 237
|||||
Qy 233 AAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLTKTVGRDFDGGQLES 292
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 GLFGLVDHSGKTGVANAGLVYFNNRLAMSEDDLPYHVKTPTGDLKTEGDFDGGQLES 297
|||||
Qy 293 TMAHPKVDPSGELFALSVDWSPKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFATTE 352
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 TMAHPKLDPSGELFALSVDYIQPYLYKFRFSGKNGEKSNDVPEPVEDPTMMHDFATTE 357
|||||
Qy 353 NFVVVPOQVVKLPKPEMLRGSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDFCFHL 412
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 NFVVVPOQVVKPMSEMRGSPVYDKNKVSREGLDKYAKDGSCLKWVEVPDQFCFHL 417
|||||
Qy 413 WNAWEEETDEVVVIGSCMTTPDSTIFNESDENLKSVLSEIRLNLTGSTRPIISNEDQ 472
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 WNAWEEETDEVVVIGSCMTTPDSTIFNECDGLKSVLSEIRLNLTGSTRPIISNEDQ 477
|||||
Qy 473 QVNLEAGMVNENMLRKTKFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPFL 532
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 478 QVNLEAGMVNENMLGRKTEYAYLALAEPPKVSFGAKVNLFTGEVEKFIYGNKYGGEPFL 537
|||||
Qy 533 FLPGE--GGEDEGYILCFVHDEKTKWSELOQVNAVSELEVEATVKLSRPVYGHGTFIG 590
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 FLPRDPSKEEDDGYILAFVHDEKWKSELOQVNAVMSLKEATVKLSRPVYGHGTFIN 597
|||||
Qy 591 ADDLAKQ 597
|||||
Db 598 ANDLANQ 604
|||||

RESULT 4
AAB72308
ID AAB72308 standard; Protein; 605 AA.
XX
AC AAB72308;
XX
DT 16-MAY-2001 (first entry)
XX
Neoxanthin cleavage enzyme-like protein amino acid sequence.
XX
Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
XX
NCP; amino acid permease; AAP; glutamic acid rich protein; GRP;
XX
pathogen resistance; abscisic acid metabolism.
XX
Lycopersicon esculentum.
XX
WO200112801-A2.
PN
22-FEB-2001.
PD
17-AUG-2000; 2000WO-US22961.
PF
18-AUG-1999; 99US-0149656.
PR
23-MAY-2000; 2000US-0206405.
XX

```


Neoxanthin cleavage enzyme; AtNCE1; abscisic acid; ABA; herbicide; stress tolerance; transgenic plant; plant breeding; antisense-therapy; plant growth protectant.

Arabidopsis thaliana.

EP1116794-A2.

18-JUL-2001.

11-JAN-2001; 2001EP-0300218.

13-JAN-2000; 2000JP-0010056.

11-JAN-2001; 2001JP-0003476.

(RIKE) RIKEN KK.

Iuchi S, Kobayashi M, Shinozaki K;

WPI; 2001-400081/43.

N-PSDB; AAD09394.

A DNA encoding a protein with a neoxanthin cleavage activity for producing transgenic plants with improved or decreased stress tolerance

Claim 3; Fig 10; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present sequence is the Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1 protein. The AtNCE1 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPR65 (CowePea Responsive to Dehydration) gene isolated from cowpea plant as a probe.

Sequence 583 AA;

Query Match 63.2%; Score 1991; DB 22; Length 583;
Best Local Similarity 64.5%; Pred. No. 1.2e-192;
Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

16 GGNHTQPLSSQSSDLSY--CSSLPMSRVTRKLNYSALHTPPALHFPKQSSNSPAIV 73
11 GGIKTWP-----QAQIDIGFRPIKQRPVIKTVQIDVTE-LTKKRQLFTPTTATP--- 62

74 VKPKAKESNTKOMLFORAAALDAEGFVSHKLEHLPLKPTADPSVQIAGNFAFVNEQ 133
63 -----QHNFLRLNIFOKARAIAIDAEALISHQSDPLPKTADPRVQIAGNFSVPDES 116

134 PVRNRLPVGKLPDSIKGVYVRNGANPLHEPVTGHFFFDGDMVHAVKFERGHSASYACRF 193
117 SVRNLVTEGTPICDGVYIRNGANPMFEETAGHLLFDGDMVHAVKIINGSASYACRF 176

194 TQNRFRQERQLGRVPFPAIGELHGTIARLMLFYARAAAGIVDPAHGTGVANAGLVY 253
177 TKTERLVQEKRLGRVPFPAIGELHGHGTIARLMLFYARGLCGLINQNGVGVANAGLVY 236

254 FNGELLAMSEDDLPYQVQITPNDGLKTVGRFPDFDQGLESTMIAPHKVDPSEGGELFALSVD 313
237 FNNLLAMSEDDLPYQVQITPNDGLKTVGRFPDFDQGLESTMIAPHKVDPSEGGELFALSVD 296

314 VSKPYLKYFRFSPDGKSPDVEIQLDQPTMMHDFAITENFVVVVDQGVVKLFPEMIRGG 373

Db 257 VVKPYLKYFRFSPDGKSPDVEIQLEPTPTMHDFAITENFVVVVDQGVVKLFPEMIRGG 356
QY 374 SPVYDKNKVARFGLLDKYAEDSSNIKWIDAPDPCFHLNNAWBPETDEVVVIGSCWTP 433
Db 357 SPVVDGKVKSRGLGIMPKDATEASQIIWVNSPETFCFHLNNAWBPETDEVVVIGSCWTP 416
QY 434 PDSIFNESDENLKSVLSEIRLNLTGSTRRTPIISNEDQOVNLEAGMVRNMLGRKTKFA 493
Db 417 ADSIFNERDESLSVLSIRLNLTGSTRRTPIISNEDQOVNLEAGMVRNMLGRKTKFA 474
QY 494 YLALAEPMKVSFGAKVDLTITTEGKVKHLYGDNRYGGEPLFLPGEGG-----BEDGYYILCF 549
Db 475 FLALAYPMKVSFGAKVDLTITTEGKVKHLYGDNRYGGEPLFLPGEGG-----BEDGYYILCF 534
QY 550 VHDEKTKWSELOIYNVNSLEVEATVKLPSRVPYGFHGTFFIGADDLAKQV 598
Db 535 VHDEETKTSELOIYNVNSLEVEATVKLPSRVPYGFHGTFFIGADDLAKQV 583

RESULT 8
AAE04788
ID: AAE04788 standard; Protein; 504 AA.
XX
AC AAE04788;
XX
DT 10-SEP-2001 (first entry)
XX
DE Zea mays neoxanthin cleavage enzyme, VP14.
XX
KW Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant.
OS Zea mays.
XX
PN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
WPI; 2001-400081/43.
DR N-PSDB; AAD09400.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is the
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1 protein.
CC The AtNCE1 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPR65 (CowePea Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.
XX

SQ Sequence 604 AA;

Query Match 61.3%; Score 1930; DB 22; Length 604;
 Best Local Similarity 63.3%; Pred. No. 2.1e-186;
 Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSOSSDLSYSSLPMSRVTRKLNVSALHTP---PALHFPKQSSNSPAIVV--- 74
 DB 17 PARSARAS-----NSVRFSRAVSSVPPAECLQAFHPKPVADLPAPSKPAALVAPGHA 71

QY 75 -KPAKESNTKOMNLFQRAAAALDA-AEGFLVS-HEKHLPLPKTADPSVOIAGNFAPVN 131
 DB 72 AAPRAEGGKQLNLFQRAAAALDAFEEGVANVLERPHGLPSTADPAVOIAGNFAPVG 131

QY 132 EOPVRRNLPPVVGKLPDSLKGVYVRNGANPLHEPTVGTGHRFFDGDGMVHAKFEHGS-A 190
 DB 132 ERPPVHELPPVSGRIPPFDGVYARNGANPCDPVAGHILFDGDGMVHALRTRNGAESYA 191

QY 191 CRFTQTNRFVQEROLGRVPFPAKIGELHGHGTGIARLMLFYARAAAGIYDPAHGTVANAG 250
 DB 192 CRFTETARLROERATIGRPVFPKAIKELHGHSGIARLALFYARAAACGLVDPSAGTVANAG 251

QY 251 LVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRFDGQLESTMIHAKPKVDPSGELFAL 310
 DB 252 LVYFNGRLAMSEDDLPYHVRVADGDLETVGRYDFDGLGCAMIAHPKLPDPAIGELHAL 311

QY 311 SYDVWSKPYLKYFRSPDGKSPDVEIQDQPTMMHDFAITENFVVVPDQOVFKLPEMI 370
 DB 312 SYDVIKRPLYKYFRPDGKSDDEVEIPEQPTMIHDFAITENFVVVPDQOVFKLQEML 371

QY 371 RGSFVYVDKKNVAFGLDKYAEEDSNIKWIDAPDCFCFHLNWAPEPTEDEVVITGSC 430
 DB 372 RGSFVYVDKKTSTRFGLVLEKHAADASEMAWVDVDFCFHLNWADEATEGEVVITGSC 431

QY 431 MTPDPSIFNESDENLKSVLSEIRLNLTGSTRPIISNEDQOVNLEAGVMNRNMLGRKT 490
 DB 432 MTPADSI NESDERLESVLTEIRLDARTGRSTRVLP-PSQENLEVGVMNRNLLGRES 490

QY 491 KFAYLALAEWPVKVSGFAKVDLTITGEVKKHLYGDNRYGGPFLFPGGGGE-----EDEGY 545
 DB 551 VLTEVHDERACTSELLVYNAADIRLEATVQLPSRVFPFGHGTFTITGQLEAQ 602

RESULT 9
 ID AAB72309
 AC AAB72309 standard; Protein; 604 AA.
 AC AAB72309;
 XX 16-MAY-2001 (first entry)
 XX Neoxanthin cleavage enzyme-like protein amino acid sequence.
 DE Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
 KW NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
 KW pathogen resistance; abscisic acid metabolism.
 XX Zea mays.
 OS Zea mays.
 XX WO200112801-A2.
 PN 22-FEB-2001.
 XX 17-AUG-2000; 2000WO-US22961.
 XX 18-AUG-1999; 9905-0149656.
 PR 23-MAY-2000; 2000US-0206405.
 XX (PION-) PIONEER HT-BRED INT INC.

PA (CURA-) CURAGEN CORP.
 XX Bidney DL, Crasta OR, Hu X, Lu G;
 XX WPI; 2001-211215/21.
 XX Novel isolated defence-related signalling gene isolated from sunflower
 encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
 acid-rich protein useful for increasing resistance of plant to a
 pathogen
 XX Disclosure; Fig 1; 135pp; English.
 XX This invention relates to defence-related signalling genes isolated from
 the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
 enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
 protein (GRP). The signalling gene is useful for increasing the
 resistance of a plant to a pathogen such as fungus, virus, bacterium,
 nematode or insect (e.g. European corn borer), preferably
Sclerotinia spp., *Phoma spp.*, or *Phomopsis spp.* by stably incorporating a
 construct containing the gene into the genome of the plant. The gene is
 useful for regulating gene expression in a plant, in response to a
 stimulus such as infection with a pathogen, damage from a pathogen,
 hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
 oxalic acid or expression of a gene encoding oxalic acid oxidase. The
 genes are also useful for stem-preferred regulation of gene expression in
 a plant. The genes are useful in agriculture, particularly in the
 breeding of crop plants with improved agronomic traits, for modifying
 abscisic acid (ABA) metabolism and for modifying amino acid transport and
 content in plants. The present sequence represents a neoxanthin cleavage
 enzyme-like protein from *Zea mays* used in the characterisation of
 sunflower NCE.
 XX Sequence 604 AA;
 QY Query Match 61.3%; Score 1930; DB 22; Length 604;
 Best Local Similarity 63.3%; Pred. No. 2.1e-186;
 Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 577 AA;

Query Match 52.8%; Score 1663.5; DB 23; Length 577;

Best Local Similarity 55.4%; Pred. No. 2.2e-159;

Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

QY 24 LSSQSSDLSYSSSLPMASRVTRKLNVS-ALHTPPALHFKQSSNPAIVVKKPAKESN 82
 Db 9 LLPTKISRSHLLPOPKVANISRLLINPKTIPLDTSVPSP-----VKLKPTYPN 62
 QY 83 TKQMLFORAAALDAAE-GFLVSGEKLHLPKLTADPSVQIAGNFAVPNEQPVRENLPV 141
 Db 63 ---LNLQKLAATLMDKTIESTSVIPMEQNRPLPKPTDPAVQLSGNFAVNECPVONGLEV 119
 QY 142 VGKLPDSIKGVYVRNGANPLHEPVVTGHHFFDGDGMHVAVK--FEHGSASYACRETQTNRF 199
 Db 120 VQIPLSCUKGYIIRGANPMFPPLAGHHLFDGDGMHVAISGFDN-QVSYSCRYTKTKRL 178
 QY 200 VQERQLRVPVPKPAIGELHGHGTGLARLMFLFVAAGAAGTVDPAHGTGVANAGLVFNGRL 259
 Db 179 VOETALGRSVFPKPIGELHGHSGGLARLALFTARAGIGLVGTRGMGVANAGVVFNGRL 238
 QY 260 AMSDDLPLQYVOITPNGDLKTVGRFEDGQLESTMIARHPKVDPSGELFALSVDVWSKPY 319
 Db 239 AMSDDLPLQYVKIDQGDLFTIGRFDDQIDSSVIAHPKYDATIGDLHTLSYNVLKKPH 298
 QY 320 LKYRFPDGTGSDVETQLDQTMHHDFAITENPVVDDQVVFVKLPFEMIRGGSPVYVD 379
 Db 299 LRYLKFNTGKKTRDVEITLPEPTMIHDFATENFVFPDQMVFKLSEMIIRGGSPVIYV 358
 QY 380 KNKVAREGILDKYADESSNLIKWIDAPDCFCFLHNAEEPTDE---VVVIGSCMPPPD 435
 Db 359 KEMARGVLSKQDLTSDINWDVDPDCFCFLHNAEE-RTGEGDPVIVVIGSCMPPPD 417
 QY 436 SFNSEDENLKSVLSEIRLNKLTGSTRPRLIISNEDQVNLFAGMVNRNMLGRKTKFAYL 495
 Db 418 TIFSEGEPIRVELSEIRLNKRTKESNKKVTVTG---VNLKAGHINRSYVGRKSQFYI 473
 QY 496 ALAEPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLPLPGEGEDEGYTLCLFVHDEK 555
 Db 474 AIADPWPKSGIAKVDIQNTGVSEFNFGPGRFGGECFVPEGEDEKGYVMGEVGRDEEK 533
 QY 556 WKSELOIVNAVSLSEVATVVKLPSPVYPYGFHGTFTIGADDLAKQV 598
 Db 534 DESEFVVVDATDMKQVAARLPERVPYGFHGTFTFSENQLEQV 576

RESULT 12

AAG31333

ID AAG31333 standard; Protein; 595 AA.

XX

AC AAG31333;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 37612.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

OS EP1033405-A2.

PN

XX 06-SEP-2000.
 PD
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.

BEST AVAILABLE COPY

```
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143242.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.7%; Score 968; DB 21; Length 595;
Best Local Similarity 36.7%; Pred. No. le-89;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVSSALHTPPALHEPKQSSNPAIVVKKAKESNTKOMNLFORAAAAADAAE 101
DB 42 SPITNPSSNDNRNPKTLH---NRTNHTLVSSPPKLRPEMTLALF---TTVEDVIN 94
QY 102 GFLVSHEKHLPLKPTADPSVQIAGNFAPYNEQPVRRNLPV-GLKLPDSIKGVVRNGANP 160
DB 95 TEIDP-----PSRPSVDPKHVLSDNFAPVLDLPPTDCEIIHGTLPLSLNGAYIRGNPN 149
QY 161 LHEPVTGHEFFDGDGMVHAKVEHGSASVACRFTOTNRFBVRQRLGRPVFPKAIHELGH 220
DB 150 QFLPRGPHLFDGDMGLHAIKTHNGKATLCRSYVTKYKYNVEKQTGAPVMPNVFSGENV 209
QY 221 T-GIARLMFYARAAAGIVDPAHGTGVANAGLVYENGRLIAMEDDLPOVQOITPNDLK 279
DB 210 TASVARGALTAARVLITGQYNPNVNGIGLANTSIAFFSNRLFALGESDLPYAVLTSBGDIE 269
QY 280 TVGREFDQOQLESTMIAPKVPDPSGELFALSVDVVKPYLKYFRFSPDGTSPDVEI-Q 338
DB 270 TIGRYDFDQKLAWSMTAHPKTDPTIGETFAKYGVP-PPFLTYFRFDSAGKKQRDVPIFS 328
QY 339 LDQPTMHDFALTENFVVVPDQQVYFK---LPEMIRGGSPVYVYDKNKFARFGLDKYAE 395
DB 329 MTSFSLDFAITKRAHFAETQLGMRNMLDLVLEGGSPVCTDNGKTPRLGVIPKYAGD 388
QY 396 SSNIKWIDAPDCFCFHLNANFEPETDEVVVGSCMTPPDSIFNDESDENLKSVLSEIKLN 455
DB 389 ESEMAWFEVPGNIITHAINAWDEDDGNSVLAPNIMIEHTLERMD-LVHALVEKVKID 447
QY 456 LKTGESIRRPDIISNEDQOVNLEAGMVNRNMLGRKTKFAYLALAEPPWPKVVSFAKVDLTG 515
```

Qy	42	SRVTRKLVNSALHTPPALHFFPKOSSNSPALVVKPKAKESNTKOMNLFORAAAAALDAAE	101
Db	42	SPITNPSONDRNRNKKPTLH---NRTNHTLVSSPPKLRPEMTLATALF---TTVEDVIN	94
Qy	102	GFLVSHKELHPLKPTADPSVQIAGNFAVNEQPVRRNLPVV-GKLPDSIKGVTVRGANP	160
Db	95	TFIDP-----PSRPSVDPKHVLNSDNFAPVDELPPDCEIITHGLPULSUNGAYIRNGPNP	149
Qy	161	LHEPVTGHHFFDGDGMVHAVKFFEHGSASVACRTQTNRVQERQLGRPFVFKAIAGELHG	220
Db	150	QFLRGRGFYHLFDGDMHLA.KIHNGKATLCSRVKYKYKYNVEKQTGAPVKNPVFSFGNV	209
Qy	221	T-GIARLMLYARAAGIVDPAGHTGVANAGLYFYNGRIAMSEDDLPTQVOITPMDGLK	279
Db	210	TASVARGALTAARVLTGQYNPVGIGLANTSLAFFSNRLFALGESDLPYAVRLTESGDTE	269
Qy	280	TVGRFEDGOLSTMIAPHPVDPESELFPALSYDVVYSKPYLKYFRFSPDGTKSPDVEI-O	338
Db	270	TIGRYDEGKLAMSMYTAHPKIDITGETFAFRYGPV-PPFLTYFREDSSAGKKQORDVPFS	328
Qy	339	LDOPTMMHDFAITENFVVVPDQGWFK---LPEMIRGSGPVVYDKNKVAREGILDKYARD	395
Db	329	MTSPSFLHDFAITKRAHIFAEIQLGMRNMMLDLVLEGGSPVGTNDCKTFLGVIFKYAGD	388
Qy	396	SSNIKWLDAPDCFCFHLHNAWEPETDEVVYIGSCWTPPDSTFNESDENLKLVLSEIRLN	455
Db	389	ESEMKNFEVPGFNI.IHAINAWDEDDGNSVVLIAPNTMSIEHTLRMD-LVHALVEKVKID	447
Qy	456	LKTGSTRRTPIINEDQQVNLBAGVNRNMLGRKTKFEYALALAEPPWKVSGFAKVDLITG	515
Db	448	LVTGIVRRHPISR-----MLDFAVLNPAFLGRCSRYVYAAIGDPMKTSIGVVKLDVSKG	502
Qy	516	E-----VKKHLGYDNRGIEPLFPGGG-----EDEGVILCFVHDEKTKWSELQIVNAV	566
Db	503	DRDDCTVARMYSGCYGGEPEFVARDPGNPAEEDDGVVYIVVHDEVTGESKFLVMDAK	562
Qy	567	S---LEVENATVKLPSRVPYGFHGTFIGADDIK	596
Db	563	SPELEIAVAALPRRPYGFHGLFVKESL NK	594

[illegible]

AAB72306
 ID AAB72306 standard; Protein; 595 AA.
 XX
 AC
 AAB72305;
 XX

XX	Neoxanthin cleavage enzyme-like protein amino acid sequence.
DE	
XX	Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme
KW	NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
KW	pathogen resistance; abscisic acid metabolism.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WC200112801-A2.
XX	
PD	22-FEB-2001.
XX	
PF	17-AUG-2000; 2000WO-US22961.
XX	
PR	18-AUG-1999; 99US-0149656.
PR	23-MAY-2000; 2000US-0206405.

PA (PION-) PIONEER HI-BRED INT INC.
PA (CURA-) CURAGEN CORP.
XX
PI Bidney DL, Crasta OR, Hu X, Lu G;
XX
WPI: 2001-211215/21
DR

Db 142 AARVITGOYNPVNGIGLANTSLAFFSNLFAIGESDLPYAVRLATESGDIETIGRYDEGK 201
QY 290 LESIMIAHPKVDPSGELFALSVDVWXPYLKYPFSPDGTKSPDVEI-QLDQPTMMHDF 348
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
Db 202 LAMSTAHPKTDPITGETFAFRYGPV-PPFLITYFRFDSAGKKQRDVPFSPFSLHDF 260
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
QY 349 AITENFVVVPDQOVVK---LPEMIRGGSPVYDKNKYAREGILDKYAEDSSNIKWIDAP 405
||| : : : : : |||| | : |||| | : |||| | : |||| | : ||||
Db 261 AITRHAIFAIOIGMRNMIDLVLGGSPVGTONGKTPRLGVTPKYAGDESEKWFVP 320
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
QY 406 DCFCFHLNAAEPEPTDEVVVIGSCMTPPDSIFNESDENLKSIVLSEIRLNLKTGESTRRP 465
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
Db 321 GFNIHAINADEDDGNSVLIAPNIMSIEHTLERMD-LVHALVEKVKIDLVITGIVRHP 379
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
QY 466 IISNEDQOVNLEAGMVRNMLGRKTKFAYLALAEFPWKVSGFAKVDLTJGE-----VKKH 520
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
Db 380 ISAR-----NLDEAVINPAELGRCRSRYVYAAIGDPMKISGVVKLDYSKGRDDDCIVARR 434
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
QY 521 LYGNRYGGEPLFLPGEGG----BEDSYILCFVHDEKTKWSEIQVNAVS--LEVEATV 574
||| : : : : : |||| | : |||| | : |||| | : |||| | : ||||
Db 435 MYGCGYGGEPFFVARDPGNPEAEEDDGYVYVYVHDEVGTGESKFLVMDAKSPELEIVAAY 494
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
QY 575 KLPSRPVYGFHGTFIGADDLAK 596
| : |||| | : || | : |||| | : |||| | : |||| | : |||| | : ||||
Db 495 RLPRRPVYGFHGLFVKESDLNK 516

Search completed: May 19, 2003, 15:47:11
Job time : 45 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:15 ; Search time 29 Seconds
(without alignments)
607.736 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASTTATAAASGRWLGNGHT.....VPYGFHGTFIGADLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	210	6.7	533	US-09-385-259-2	Sequence 2, Appli
2	210	6.7	533	US-09-645-370-2	Sequence 2, Appli
3	209	6.6	533	US-08-488-305A-6	Sequence 6, Appli
4	123.5	3.9	1043	US-08-928-361B-30	Sequence 30, Appli
5	123.5	3.9	1721	US-08-700-651-5	Sequence 5, Appli
6	123.5	3.9	1721	US-08-928-361B-6	Sequence 6, Appli
7	116.5	3.7	1042	US-08-928-361B-11	Sequence 11, Appli
8	116.5	3.7	1837	US-08-928-361B-5	Sequence 5, Appli
9	106	3.4	412	US-09-461-697-7	Sequence 7, Appli
10	104.5	3.3	2172	US-08-611-107-31	Sequence 31, Appli
11	98	3.1	741	US-09-001-984C-106	Sequence 106, App
12	97.5	3.1	10182	US-09-134-001C-3159	Sequence 3159, Ap
13	97	3.1	516	US-08-097-829-4	Sequence 4, Appli
14	97	3.1	516	US-08-577-403-4	Sequence 4, Appli
15	96	3.0	1410	US-09-335-409-3	Sequence 3, Appli
16	96	3.0	1410	US-09-568-102-3	Sequence 3, Appli
17	96	3.0	1410	US-09-567-969-3	Sequence 3, Appli
18	96	3.0	1410	US-09-568-480-3	Sequence 3, Appli
19	96	3.0	1410	US-09-568-486-3	Sequence 3, Appli
20	96	3.0	1410	US-09-568-472-3	Sequence 3, Appli
21	96	3.0	1410	US-09-567-899-3	Sequence 3, Appli
22	95.5	3.0	1802	US-09-322-478-18	Sequence 18, Appli
23	94.5	3.0	620	US-09-000-145-5	Sequence 5, Appli
24	94	3.0	1471	US-08-755-188	Sequence 188, App
25	92.5	2.9	868	US-08-907-166-2	Sequence 2, Appli
26	92.5	2.9	1057	US-09-697-367-23	Sequence 23, Appli
27	92	2.9	205	US-09-385-259-3	Sequence 3, Appli

28	92	2.9	205	4	US-09-645-370-3	Sequence 3, Appli
29	91.5	2.9	230	4	US-09-625-188-18	Sequence 18, Appli
30	91	2.9	733	3	US-08-725-459B-22	Sequence 22, Appli
31	90.5	2.9	470	1	US-08-090-523-10	Sequence 10, Appli
32	90.5	2.9	470	1	US-08-398-627-10	Sequence 10, Appli
33	90.5	2.9	470	1	US-08-406-858-10	Sequence 10, Appli
34	90.5	2.9	470	1	PCT-US91-04036-10	Sequence 10, Appli
35	90.5	2.9	470	5	PCT-US94-05275-10	Sequence 10, Appli
36	90.5	2.9	1038	4	US-08-334-179A-2	Sequence 2, Appli
37	90	2.9	390	4	US-08-861-774E-18	Sequence 18, Appli
38	90	2.9	844	3	US-09-029-267-20	Sequence 20, Appli
39	90	2.9	1253	1	US-08-252-966B-12	Sequence 12, Appli
40	90	2.9	1261	1	US-08-252-966B-18	Sequence 18, Appli
41	88.5	2.8	383	4	US-09-530-212A-2	Sequence 2, Appli
42	88.5	2.8	462	1	US-08-548-509-5	Sequence 5, Appli
43	88.5	2.8	872	2	US-08-844-057-2	Sequence 2, Appli
44	88.5	2.8	872	4	US-09-006-730-2	Sequence 2, Appli
45	88	2.8	639	4	US-09-347-801-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-385-259-2
; Sequence 2, Application US/09385259
; Patent No. 6201114
; GENERAL INFORMATION:
; APPLICANT: Aguirre, Gustavo D.
; APPLICANT: Aguirre, Gregory M.
; APPLICANT: Ray, Kunal
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/385,259
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 60/103,219
; EARLIER FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-385-259-2

Query Match	6.7%	Score 210;	DB 4;	Length 533;
Best Local Similarity	19.9%	Pred. No. 5.4e-13;		
Matches 113;	Conservative 98;	Mismatches 218;	Indels 140;	Gaps 25;
Qy	120	SVOI---AGNFAPVNEQPVRRNLP-----VVGKLPDSIKGVYVRNGANPLH---EPVTGHH	169	
Db	2	SIQVEHPAGGYKKLFEIWEELSSPLTAHTVGRIPMLTGLSLLRCGPGLFVSGSEPF--YH	59	
Qy	170	FDGDMGHVAVKHEGHSYACRTQTRFVQE-----RQLGRPVFPKAGLGHGT	221	
Db	60	LFDGQALLHKDFDKGHHYTHRRFTDTAYVRAMTEKRVITFETGTCAPDPCKNI----	115	
Qy	222	GIARLMFYARAAGIVDPAHGTGVANAGLVYFNGRLLAMSED-----DLPVQVQITPNG	276	
Db	116	-FSRFFSYF-----RGVEYTDNALV-----NYVPVEDYYACTETFTIKINPE-	158	
Qy	277	DLKTVGRFDFDGOLE-STMIAPKVPDES-----GELFALSVDVVSPLYKLYFRFS	326	
Db	159	TLETIKQVDCNYSVNGATAHPIENDGTVYVNICNCFGNFSTAYNIVKIPPLQADKED	218	
Qy	327	PDGTSKSPDEIQLD-----QPTMMHDAITENFVVPDQVQVFKLPEMIRGS-----	374	
Db	219	P--ISKSEVVQVPCSDRFKESYVHSEGLTPNYIVFVETPVKINLLKPLSSWLGWANYM	276	
Qy	375	-----PVVYDKNKVARFGILDKYAEKSSNKKIDWADPCFCFHLNNAWEEPEDEV	424	
Db	277	DCFESNETMGVWLHIADKKRKKYLNKYRTSSFNL-----FHHINTYEDNEF--L	324	

QY 425 VVIGSCMTTPDSTFNSD-ENLKSVLSEIRLNKLTGSTRRPILISNEDQQVNLEAGWNR 483
Db 325 IVDLCCKWGFEEFYNYLYLANRENWEEVK---KNARKAPQPEVRRSVPLNIDKADTGK 381
QY 484 NMLGRKTKFAYLAL-----AEPWPKVSG-----FAKVDLTT--GEVKKHLYG-----523
Db 382 NLVTLPTNTATATLRSDETIWLPEVLFSGPROAFEPQINYQYKYGKPYTAYAGLNLH 441
QY 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTKW--SE 559
Db 442 FVDRCKLNKVTKETWVQEPDSYSEPIFVSHPDALDEDDGVLVSVVSPGAGOKPAY 501
QY 560 LQTVNAVSLVEATVKLPSRPVPGFHGT 588
Db 502 LLILNAKDLSEVARAEVINIPVTFHGLP 530
RESULT 2
US-09-645-370-2
; Sequence 2, Application US/09645370
; Patent No. 6428958
; GENERAL INFORMATION:
; APPLICANT: Aguirre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; TITLE OF INVENTION: IN DOGS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/645,370
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/385,259
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: 60/103,219
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-645-370-2
Query Match 6.7%; Score 210; DB 4; Length 533;
Best Local Similarity 19.9%; Pred. No. 5.4e-13;
Matches 113; Conservative 98; Mismatches 218; Indels 140; Gaps 25;
QY 120 SVQI---AGNFAFVNEQPVRRNLP---VVGKLPDSTKGVYVRNGANPLH---EPVTGHH 169
Db 2 SIQVEHPAGYKKLFETVELSPLTAHVTRGRIPLWLTGSLRCGGLFVGVSEPF--YH 59
QY 170 FPDGGMVHAVKFEHGSASYACRFTQTNREVOE-----RQLGRPVFPKATGELHGT 221
Db 60 LFDGQALLHKFDEKEGHVYHRRFIRTDAYVRAMTEKRIVITFETGCAFPDCKNI---115
QY 222 GIARLMLFYARAAGIVDPAGHCTGVANAGLYFNGRLAMSED-----DLPYQVQITPNG 276
Db 116 -FSRFSYF-----RGVEVTDNALV---NVPVGEDYACTETNFTIKNPE-158
QY 277 DLKTVGRFDPDGOLE-STMLAHKVPDPES-----GELFALSVDVSKPVLYKVPRES 326
Db 159 TLETIKQVDCNIVSVNGATAHPIENDGVYVINGNFCGKNFSIAYNIVKIPLOADKED 218
QY 327 PDGKSPDVEIQLD-----OPTMMHDAITENFVVVDPQVFKLPEMIRGS-----374
Db 219 P--ISKSEVVVQPCSDRRKPSVYHSGPLPNVIVFETPVKINLLKFLSSWLWGANYM 276
QY 375 -----PVVYDKNKVARFGTLDKXAESSNKKWIDAPDFCFELWNAWEPETDEV 424
Db 277 DCFESNETMGWLHIADKKRKKVYLNKKYRTSSFNL-----PFIINTYEDNEF--L 324
QY 425 VVIGSCMTTPDSTFNSD-ENLKSVLSEIRLNKLTGSTRRPILISNEDQQVNLEAGWNR 483

Db 325 IVDLCCKWGFEEFYNYLYLANRENWEEVK---KNARKAPQPEVRRSVPLNIDKADTGK 381
QY 484 NMLGRKTKFAYLAL-----AEPWPKVSG-----FAKVDLTT--GEVKKHLYG-----523
Db 382 NLVTLPTNTATATLRSDETIWLPEVLFSGPROAFEPQINYQYKYGKPYTAYAGLNLH 441
QY 524 -----DNRYGGEPLFLP-GEGGEDEGYILCFVHDEKTKW--SE 559
Db 442 FVDRCKLNKVTKETWVQEPDSYSEPIFVSHPDALDEDDGVLVSVVSPGAGOKPAY 501
QY 560 LQTVNAVSLVEATVKLPSRPVPGFHGT 588
Db 502 LLILNAKDLSEVARAEVINIPVTFHGLP 530
RESULT 3
US-08-488-305A-6
; Sequence 6, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof; Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vineet
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-488-305A-6
Query Match 6.6%; Score 209; DB 1; Length 533;
Best Local Similarity 20.7%; Pred. No. 6.8e-13;
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;
QY 105 VSHEKULHPLPKTADPSVQIAGNEAPVNEQPVRRNLP---VVGKLPDSIKGVYVRNGANP 160
Db 1 MSSQVEHP-----AGYKKLFETVELSSPLTAHVTRGRIPLWLTGSLRCGGL 49
QY 161 LH---EPVTGHHFDPDGMVHAVKFEHGSASYACRFTQTNREVOE-----RQLGRPV 209
Db 50 FEVCSEPF--YHLDGQALLHKFDEKEGHVYHRRFIRTDAYVRAMTEKRIVITFETGCA 107
QY 210 FPKAIGELGHGTGIARLMLFYARAAGIVDPAGHCTGVANAGLYFNGRLAMSED-----264
Db 108 FPDPCKNI-----FSRFSYF-----RGVEVTDNALV---NVPVGEDYACT 147
QY 265 DLPYQVQITPNGDLKTVGRFDPDGOLE-STMLAHKVPDPES-----GELFALSVDV 314

Db 148 ETNFKVNEP-TLETKQVLDLCNYSVNGATAHPHIENDGTVYNIGNCFKNEFSIAYNI 206
Qy 315 VSKPYLYFFSPDGTSPQVEIQD-----QPTMHDFAITENFVVPVQVVKLPPEM 369
Db 207 VKIPPLADKEDP--ISKSEIVVQPCSDREKPSYVHSFGLTPNYIVFVETPVKINLFKE 264
Qy 370 IRGSPVYVYDKNKVARF-----GILDKYAEDSSNIKWIDAPOCFPHLWANEHP 419
Db 265 LSSWS--LWGANYMDCFESNETMGWLHIAKDKKKYKINNYRTSP-FNLFHHINTYEDH 321
Qy 420 ETDEWVIGSOMTPDPSIFNESD-ENLKSVLSEIRLNKLTGSTRRPILISNEDDOVNEA 478
Db 322 EF--LIVDLCCWKGFVYVILYLANLRENWEVK---KNARKAPQEVRYVLPNLIDK 376
Qy 479 GVMNRNM--LGRKTKFAYALAEF-W--PKY--SG-----FAKVLDLT--GEVKKHLYG 523
Db 377 ADTCNLTLPNTTATATILCSDETINLEPEVLFSGPRGAEFPQINQYKYGKPYTYAYG 436
Qy 524 -----DNRYGGELFLP-EGGGEDEGYILCFVHDEKTV 556
Db 437 LGLNHFVPDRCLKNVKTKETWVQEPDSYSEPIFVSHPDALBEDDGVLSVYVSPGAG 496
Qy 557 K--SELOTVNAVSEVEATVKLPSRVYGPHTF 588
Db 497 QKPAYLLILNAKDLSEVARAEVNIPIVTFHGLF 530

RESULT 4

US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928.361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026.062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Veruy, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-30

Query Match 3.9%; Score 123.5; DB 3; Length 1043;
Best Local Similarity 20.6%; Pred. No. 0.0021;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;
Qy 5 TATAAAGCRWLGGNHTQPLPSSOSSDLISYCSLL--PMASRVTRK-----LNVSSALHTPP 58
Db 153 TIAGIVSG-----ISASESL-LSQKSALIDATNNVVGEEFGLLNPAIGVMIPG 200
Qy 59 ALHFPKSSNSPAI-----VVKPKAKESNTKOMNL-----FORAAALD 98
Db 201 FLGSEQIOFSPEDGGIIPPEVAANADKFLSIPSPVESIPEKDOKIDSISEMYD 260
Qy 99 AAGFLVSHKHLPLPKTADSPVOIAGFAPVNEQPVRRNLPPVVGKLPDSIKGVVRNGA 158
Db 261 IESGRLLGQVSKRPPIGS-----IAGLNPIMKTIPTQD-SVTGKPIDPTIGL----- 307
Qy 159 NPLHEPVTHGHFFDDGDMHVAHAKFEHGSASY--ACRFTQTNRFFVQEROLGRFPVFKAI 216
Db 308 -PFNPP-TGH-----LINPTNNNTMDSSEAGAYKAYAVSNGIKTDNVYGLPV-----GE 353
Qy 217 LHGTGTARLMLFYARAAAGIVDPAHGTGVAN--AGLVYFNGRL-LAMSEDDLPQVOIT 273
Db 354 ITGLPKDPGSDIPFNSTTGLVDPSTGKPINNSTAGIV--SGKPLGPPIEDE----- 403
Qy 274 PNGDLKIVGRED-----FDGQLESTMIAPKVPDPESEGELEFALSYDVVSKPKLYFRFSP 327
Db 404 -NGNL-----FDPSTNLPIDGN-----NOLVNPETNSTVSGSTSGTTP-----KP 443
Qy 328 -----DGTKSPDVEI--QLDQ-----PTMHDFAITENFVVPVQVVF 364
Db 444 GIPVNGGVVPEEAKQADKQKGLIVPTNINSKDPVTNTQYNSNTTGNIIINPETGKV- 502
Qy 365 KLPENIRG-----GSPVYDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLWANE 417
Db 503 -IPGSLGPLNAPSFNTTQQQDE-----ITGKPDVTVTGLPY----- 538
Qy 418 EPETDEVVVGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRPILISNEDDOVNE 477
Db 539 DEPTGEIIDPAIKLPIGVSAGD-----EILTEV-LAITTDEVTVGLFI-----DLE 583
Qy 478 AGMVNRNMLGRKIKFAYLALAEPPKVSFGAKV---DLTTGEVKKHLYGDNRYGGEPLFL 534
Db 584 TGL-----PRDPVSGLPOLPNGTLVDPNSKKKPIPGS--HSG---FI 619
Qy 535 PEGGEE 541
Db 620 NGTSGEQ 626

RESULT 5

US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM
; FILE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700.651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum

RESULT 7

US-08-928-361B-11
 ; Sequence 11, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; TITLE OF INVENTION: SPECIES INFECTIONS
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Verny, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1042 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-928-361B-11

Query Match 3.7%; Score 116.5; DB 3; Length 1042;
 Best Local Similarity 19.9%; Pred. No. 0.011;
 Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHPFKQSSNSPAI-----VVKPKAKESNTKOMNL----- 88
 Db 190 LNPATGVMTIPGSLGSPSEQTPFSEIDGGIIPPEVAANADKFKLSIPSPVSEIPEKDO 249
 QY 89 -PQRAAALDAEGLVSHKHLPLKPTADPSVQIAGNFAVPNEQVRNLPVVGKLPD 147
 Db 250 KIDSISELMYDIESGLIGOVSKRPIPGS-----IAGSLNFMKPTPTD-SVUGKPID 302
 QY 148 STKGVVVRNGANPLHEPTVGHFFDGDGMVHAKVEHGSASY--ACRFTQTRFVQEROL 205
 Db 303 PTTGL-----PFNPP-TGH-----LINPTNNTMDSFAGAYKYAVSNGIKTDNVY 347
 QY 206 GRPV-----FPKAIAGELHGTGIAIPLMFLYARAAAGIVDPAGHTGVAN-- 248
 Db 348 GLPVDIITGLPKDPVSDIPFNSTTGL-----VDFSTGKPINNYT 387
 QY 249 AGLVFNGLIAMSDDIPYQVITPFGDL---KTVGRFDFDQLESTMIAPKVDPESG 305
 Db 388 AGIV-SGRGLPPIEDE-----NGNLPDPSTKLPIDGNNQL-----VAPETN 428
 QY 306 ELFALSIVQVSKPKLYKFRFSP-----DGTKSPDVEI--QIDO----- 342
 Db 429 STVSGTSGSTKP-----KPGIPVNGGVVDPDEAKDQADKDGDLIVPTPNTSINKDP 481

QY 343 TMMHDFAITENFVVVQOVVEKLPFEMIRG-----GSPVVYDKNVKVARFGLDKYAE 395
 Db 482 VTTOYSNTTIGNINPETGKV--IPGSLPGSLNYPSENTPQOTDE-----ITGKPVDT 532
 QY 396 SSNIKWDADPCFCFHLNNAWEPEPDETVVVGSCMTTPPDSIFNESDENLKSVLSEIRLN 455
 Db 533 VTGLPY-----DPSTGEIIDPATKLPDIPGVSAGD-----EILTEV-LN 569
 QY 456 LKGTGSTRRLIISNEDQOVNLEAGMVRNMLGRKTKAYALAEPPKPVSGFAKV----DL 512
 Db 570 ITDDEVITGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 602
 QY 513 TTGEVKKHLYDNRYGGEPLFLPGEEGEE 541
 Db 603 VDPSENKKPIPGS--HSG---FINGTSGEQ 626

RESULT 8

US-08-928-361B-5
 ; Sequence 5, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; TITLE OF INVENTION: SPECIES INFECTIONS
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Verny, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1837 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-928-361B-5

Query Match 3.7%; Score 116.5; DB 3; Length 1837;
 Best Local Similarity 19.9%; Pred. No. 0.03;
 Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHPFKQSSNSPAI-----VVKPKAKESNTKOMNL----- 88
 Db 985 LNPATGVMTIPGSLGSPSEQTPFSEIDGGIIPPEVAANADKFKLSIPSPVSEIPEKDO 1044
 QY 89 -PQRAAALDAEGLVSHKHLPLKPTADPSVQIAGNFAVPNEQVRNLPVVGKLPD 147

Db 1045 KIDSISELMYDIESGRILGOVSKRPIGS-----IAGDLNPMKTTQTD-SWTGKPID 1097
QY 148 SIKGVYVYRANGANLHEPVTCGHHFPGDCMVHAKVFEHGSASY--ACRFTOTNRVQEROL 205
Db 1098 PTTGL-----FNPBP-TGH-----LINPTNNMTDSFAGAYKYAVSNGIKIDNVY 1142
QY 206 GRPV-----BPKAIGELHGHGTGIARLMLFYARAAAGVDPHAGTGVAN-- 248
Db 1143 GLPVDETGLPKDPVSDIPNSITGEL-----VDSTGKPINNYT 1182
QY 249 AGLVYFNGRLLANSEDDLPVQVITPNDL---KTVGRFDFDQOLESTMIAPKVDPESG 305
Db 1183 AGIV-SGKRLGPPIEDE-----NGNLFDPSTKLPIDGNL-----VNPETN 1223
QY 306 ELEFALSVDVVKPKYKVFRES-----DGTKSPDVEI--OLDO-----P 342
Db 1224 STVSGSTSGTKP-----KPGIPVNGGVDPDEAKQADKGLIIVPTPNSINKDP 1276
QY 343 TMMHDAITENFVVVDQVQVFKLPEMIRG-----GSPVYDKNKKVAREFGILDKYAE 395
Db 1277 VNTQYSNTITGINPETGKV--IPGSLPGSLNKPSEFNTQOTDE-----ITGRPVDT 1327
QY 396 SSNIKWIDAPDCFEHLNWAWEPEDEVVVIGSCMTPPDSIFNESDENLKSVLSEIRLN 455
Db 1328 VTGLPY-----DFTGEIIDPATKLPVPGSVAGD-----EILTEV-LN 1364
QY 456 LKGTGSTRRPIISNEDQVQVNLKAGVMNRNMLGRKTRKFAYLALAEPPKVSQFAKV--DL 512
Db 1365 ITDVEVTGLPI-----DLETL-----PRDPVSLGQLPGLNGTL 1397
QY 513 TTCEVKKHLYGDNRYGEPFLPGEGEE 541
Db 1398 VDPNKKPIGS--HSG---FINGTSGEQ 1421
RESULT 9
US-09-461-697-77
; Sequence 77, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-77
Query Match 3.4%; Score 106; DB 4; Length 412;
Best Local Similarity 20.6%; Pred. No. 0.027;
Matches 74; Conservative 45; Mismatches 131; Indels 110; Gaps 16;
QY 23 PLSSQSSDLSSYCSLLPMASVTRKLVNSALHTPPALHPKSSNSPALVVKPKAK--- 79
Db 9 PISSQTSNRSSP-DLLPREFRLVE-----VHDP-LHQPANKPKPTMLDIPSEPS 59
QY 80 -----ESNTKONMLFORAAALDAAGFLVSEKHLPLPKTADPSVQIAGNAPV 130
Db 60 LATHIQLIQHNRRLNRLIATAQAQNOQOTEG--VKTESEPLP-SCPGSPPLPDDLLEPL 116

QY 131 N-EQPVARNLPVVGKLPDPSIKAGVYVRNGANPLHEPV-----GHHF 170
Db 117 DCKNP---NAPFOIRHSDPESDFYRGKG-----EPVTELSWHSCRQLLYQAVAILAHAG 168
QY 171 FDGDMVHAKVFEHGSASYACRFTOTNRVQERQ--LGRPVFPKAI GELHGHGTGIARLML 228
Db 169 FDCANESVLETLIDVAHEYCLKFTKLLRFADREARGQTFFPDVMEQVFEHVGIGSVLS 228
QY 229 F-----YARAAAGVDPHAGTGVANAGLVVFNGRLLAMSE 263
Db 229 LQKFWOHRKIDYHMYLQISKQLSEEVER---LYNEKAT-----E 266
QY 264 DDLPYQVOITPNDGLKTVGRFDFDQOLESTMIAPKVDP-----ESGELFALSVDVVKSP 318
Db 267 DAKPVKIKKEEFVSDI---TFPVSEELADLASDQSLPMGVILGAQSERFPNSLEVEASP 322
RESULT 10
US-08-611-107-31
; Sequence 31, Application US/08611107
; Patent No. 5801233
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING
; TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,107
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/09340
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 08/422,560
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:221
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-611-107-31
Query Match 3.3%; Score 104.5; DB 1; Length 2172;
Best Local Similarity 19.9%; Pred. No. 0.74;
Matches 118; Conservative 80; Mismatches 247; Indels 147; Gaps 27;

QY 66 SSNSPAIVKPKAKESNTKOMNLFORAAAAADRAEGFLVSHKHLPLPKTADPSVQIAG 125
Db 1406 TGNCTVDIIRVEDPNIHK--LFYRSATPAGLHG--IALHEPKPL-----DAIDLKR 1457
QY 126 NFAPVNEQVVRNLPVVGKL-----PDSIKGVYVRNGANPLHBPVTGHHFFDGDGM--- 176
Db 1458 AARKNETTYCYDFPLAFETALKKSWESSGISHVAESNEHNGRYAEVTELIADTSNGWT 1517
QY 177 -VHAVKEHGSASVACRFTQTNRFVQERQLGRPVF-----PKAIGELHGH 221
Db 1518 PLVPERPPGSGNFGVAVANMKLSTPEPPGGREIIVVANDVTFKAGSFGPREDAFFAVT 1577
QY 222 GIA---RLMLFYARAAGIVDPAGHGTGVANAGLVFNGRLLAMSEDDLP-----YQVQITP 274
Db 1578 NLACERKIPLIYLSATAGA-----RLGVAEIKACFH---VGVSDOSPERGFHYIYTE 1629
QY 275 NGDLKTVGRDFDQQLSTMAHPKVDPESGELFALSVDVYSK----- 317
Db 1630 Q-----DY-SRLSSVIAHELKVPESGETRWVVDIVGKEDGLGCENLHSGGAIAS 1679
QY 318 PYLKYER--FSPDGTGKSPDVEI-----QLDQPTMHDFALTENFVVVPPQOV 362
Db 1680 AYSKAYRETTLTETVGRGIGAYLARGMRCIORLDQPLITGYSALNKLL--GREV 1736
QY 363 VFKLPEMIRGSPVYDKNKVARFGILDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPTD 422
Db 1737 YSSOMQL--GGPKMATNGVVHLTVSDDLGVSAILKWL-----SVVPEYVG 1781
QY 423 EVVVGSCMTPPDSTIFNESDENLKSVLSEIRLNKLTGSTRRP--IISNEDQOVNLEAGMVN 482
Db 1782 GLPLFVKSLDPPERAVTYFPE-----SCDARAAICGIDTQGWKLSGMEF 1827
QY 483 RNMLGRKTKAYLALAPWPVK--VSGPAKVDLTGTGEVKKHLYGDNRYGGGELFLPG---- 536
Db 1828 RE-----SFVTELGWAKTVITGRASAKTAQA---LLDFNR--EEDPLFILANWRG 1875
QY 537 -EGGED--PGYI---LCFVHDEKTKSE--LQIVNAVSLSEVATVKLPSPRV 580
Db 1876 FSGQRDLFEGILQAGAMVENLNTYKQPAFYIIPKAGELRGGAWVVVDOSKI 1927

RESULT 11

US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331

GENERAL INFORMATION:

; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Bellisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001,984C
; CURRENT FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 106
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-09-001-984C-106

Query Match 3.1%; Score 98; DB 4; Length 741;
Best Local Similarity 21.7%; Pred. No. 0.53;
Matches 139; Conservative 79; Mismatches 216; Indels 208; Gaps 37;

QY 97 LDRAEGFLVSHKHLPLP-----KTADPSVQIAGNFAFVNEQPV---RRNLPPV---GK 144
Db 78 MDAYRQELPIETGYLLPEPDDFTTITSGVDAEITTTAGPQLVWPVNLNARFALNANAWGS 137
QY 145 LPDSIKGVYV---RNGA--NPLPEPVTG-----HFFDGDGKVHAKVEHGSASVACR 192

Db 138 LYDALYGTVDIPEIDGAEGKPTYNKVRGDKVIAYARKFLD-----DSVPLSSGSGDATG 192
QY 193 FTOQNRFRVQERQLGPRVPPKAI-----CELHGHTGIAR-----LMLFYARAAGIVDPAH 242
Db 193 FT-----VQDGLVVALPDKSTGLANPGQFAGYTGAAESPTSVLLINHLGHEILIDPES 247
QY 243 GTGVAN-AG-----LVFNGRLAMSEDD--LPYQVQITPN-GOLKTV----- 281
Db 248 QVGTIDRAGVRDVTILESAITTMDPESVAAVDAADKVLGYRNWGLNKGDLAAAADKDG 307
QY 282 -----GREDFDQ-----LESTMAHPKVDPESGELFALSVDVYS 316
Db 308 TAFRLVLRDRNYTAPGGGQFTLPGRSLMFVRNVGHLMNDIAIVTDCGSEVFEGLMDALF 367
QY 317 KPYLYKFRFSPDGTGKSPDV-----ETQLDQPTMMH---DFAIT-ENFVVVPPQOV 362
Db 368 TGLI-----AIHGLKASDVNGPLNSRTGSIYVKP-KMHGPAEVAFTCELSRVED--- 418
QY 363 VFKLPEMIRGSPVYDKNKVARFGILDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPTD 422
Db 419 VLGLPQ-----NTMKIGIMDEERTTVNLK-----ACIKAAAD 451
QY 423 EVVVGSCMTPPDSTIFNESDENLKSVLSEIRLNKLTGSTRRP-----IISNEDQOVN 475
Db 452 RVFTINTGF-----LDRTGDEIHTSMEAGPMWRKTKMSQFWILAYEDHNDV 498
QY 476 --LEAGMVNRNMLGR-----KTKFAY--LALAPW-PKVSG-----FAKV 510
Db 499 AGLAAGFSGRAQVKGWMTMTLMADMVETKIAQPRAGASTAWVPSPATAATLHALHYHQV 558
QY 511 DLTTGTEVKKHLYGDNRYGGEPL-----FLPGEGBEDE-----GYILCFVUDE 553
Db 559 DVAA--VQDGLAGRRRTIEQLLTIPIAKELAWAPEIREEDVNNCOSILGYVVRWV-DQ 615
QY 554 KTWKSELQIVNAVSL-EVEATVKLPSPRVYGF--HGTFIGAD 592
Db 616 GVGCKVPDIHDVALMEDRATLRISQLLANLWLRHGVITSAD 657

RESULT 12

US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3159

Query Match 3.1%; Score 97.5; DB 4; Length 10182;
Best Local Similarity 19.4%; Pred. No. 61;
Matches 106; Conservative 72; Mismatches 204; Indels 165; Gaps 26;

QY 13 RWLGGNHTQFPPLSSQSDLSYCSLPMASRVTRKL-----NVSSALHTPPALHFFKQS 66
Db 673 RWOGAN-----AMNAYQIEELFSQENLQNAARSGRPIQFLVGFVEDSHHNPETL-LFVN- 726
QY 67 SNSPAIVVKKAK-----ESNTKQMNLFORAAAAADRAEGFLVSHKHLPLPKT 116
Db 727 -----LVVKKPELKHTEIYLDHNEKQDKRKFVSVKRAGHVFQVMSGLTHTVSGILPYQ 781

```
QY 117 ADFSVQIAGNAPVNEQVVRNLPVVGKLPDSI---KGVYVRNGANLHPHVPVTHGHEFDG 173
Db 117 ADFSVQIAGNAPVNEQVVRNLPVVGKLPDSI---KGVYVRNGANLHPHVPVTHGHEFDG 173
QY 782 QETRIKLTN-----EPKDSWSITGYNTLTQLNAVGTNNATKKNLALYCH---IDP 833
Db 782 QETRIKLTN-----EPKDSWSITGYNTLTQLNAVGTNNATKKNLALYCH---IDP 833
QY 174 DGMVHAKFPHGSYACRFTQTRFVQERQOLGRPVFPKAI---GELHGH-----T 221
Db 174 DGMVHAKFPHGSYACRFTQTRFVQERQOLGRPVFPKAI---GELHGH-----T 221
QY 834 GNVFITVKFGDKVEQFEIR-----SKPTTPRIITANELRGNPNHKPEIRVT 880
Db 834 GNVFITVKFGDKVEQFEIR-----SKPTTPRIITANELRGNPNHKPEIRVT 880
QY 222 GIARLMLFYARAAAGIVD-----PAHGTGVANAGLVVF-----NGRLIAME 263
Db 222 GIARLMLFYARAAAGIVD-----PAHGTGVANAGLVVF-----NGRLIAME 263
QY 881 DIENDITAKIKLVGGTGDGHDEINPYTVPENYTVVAEA---YHNDPKNGVLIIFRSS 937
Db 881 DIENDITAKIKLVGGTGDGHDEINPYTVPENYTVVAEA---YHNDPKNGVLIIFRSS 937
QY 264 D---DLPOVQVITPNSDLKTVGRF-----DFDQLE-STMIAPKVDPSGELFALSVD 313
Db 264 D---DLPOVQVITPNSDLKTVGRF-----DFDQLE-STMIAPKVDPSGELFALSVD 313
QY 938 DYLKDLPL-----SGELKAIVYVQVQNSVFSFSDTTPPTINEPAGLVH----- 986
Db 938 DYLKDLPL-----SGELKAIVYVQVQNSVFSFSDTTPPTINEPAGLVH----- 986
QY 314 VSKPYLKYFR-----FSPDGTSPDYELQDQTMHDFALTENFVVPDQV 362
Db 314 VSKPYLKYFR-----FSPDGTSPDYELQDQTMHDFALTENFVVPDQV 362
QY 987 -----KYRGDHVEITLPTONTGGSLRDVNVNLPQ-----CWTKFTIINPNNT 1032
Db 987 -----KYRGDHVEITLPTONTGGSLRDVNVNLPQ-----CWTKFTIINPNNT 1032
QY 363 VFKLPEMIRGGSPVYOKNKAARFGLDKVAEDSSNIKWIDAPDCFCFHLWNAKEEPT- 421
Db 363 VFKLPEMIRGGSPVYOKNKAARFGLDKVAEDSSNIKWIDAPDCFCFHLWNAKEEPT- 421
QY 1033 EGIL--KLIGNISNEAYNTYHENT--ATNSG-----NITNPAKIF 1072
Db 1033 EGIL--KLIGNISNEAYNTYHENT--ATNSG-----NITNPAKIF 1072
QY 422 -----DEVVITG-----SMTPTDPSIFNESDENLKVSEIRLNKLTGSTREPII 467
Db 422 -----DEVVITG-----SMTPTDPSIFNESDENLKVSEIRLNKLTGSTREPII 467
QY 1073 ILNVGKLADLNPVGLSRDQLVLTDPDSSLNSREBEVKRKISEANANIRSYLLQNPIL 1132
Db 1073 ILNVGKLADLNPVGLSRDQLVLTDPDSSLNSREBEVKRKISEANANIRSYLLQNPIL 1132
QY 468 SNEQVQV 474
Db 468 SNEQVQV 474
QY 1133 AGVNDV 1139
Db 1133 AGVNDV 1139
RESULT 13
US-09-758-829-4
; Sequence 4, Application US/08097829
; Patent No. 5498831
; GENERAL INFORMATION:
; APPLICANT: Burgess, Diane G.
; APPLICANT: Dooner, Hugo K.
; TITLE OF INVENTION: Pea ADP-Glucose Pyrophosphorylase Subunit
; TITLE OF INVENTION: Genes and Their Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,829
; FILING DATE: 19930723
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 12176-35
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-758-829-4
```

```
Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;
QY 1 MASFTATAAAYSGRWLGNGHTQPPSSSSQSSDLSSCCSLPMSRVTRKLNYSVA-----L 54
Db 1 MASMAAIGVLK-----VPPSSSSSSS-----SSSSSKAIAARNLSITSSQLCGDKI 46
QY 55 HTPPALHFPKQSSNSPAIVVVKPKA-KESNTKQNMFLFORAAALDAAEGFLV---SHEKL 110
Db 55 HTPPALHFPKQSSNSPAIVVVKPKA-KESNTKQNMFLFORAAALDAAEGFLV---SHEKL 110
QY 47 FT---VSGTTRSSGRNPFIIVSKAVSDSKNSQTCCLDPDASRVL-----GIILGGGAGTRL 99
Db 47 FT---VSGTTRSSGRNPFIIVSKAVSDSKNSQTCCLDPDASRVL-----GIILGGGAGTRL 99
QY 111 HPL-PKTAADSVQIAGNAPVNEQVVRNLPVVGKLPDSIKGVYVRNGANLHPHVPVTHG 169
Db 111 HPL-PKTAADSVQIAGNAPVNEQVVRNLPVVGKLPDSIKGVYVRNGANLHPHVPVTHG 169
QY 100 YPLTKKRAKPAVLGANRLI-----DIPVSNCLNSNLSIKVLTQFN---SASLNRR 149
Db 100 YPLTKKRAKPAVLGANRLI-----DIPVSNCLNSNLSIKVLTQFN---SASLNRR 149
QY 170 F-----FDGDMGHVAVKFE-----HGSA-----SYACRETQTN----- 197
Db 170 F-----FDGDMGHVAVKFE-----HGSA-----SYACRETQTN----- 197
QY 150 LSRAYASNLGGYKNEGFEVLAAQOSPENPNWFGQTADAVRQYLWLFEEHNVLVLAG 209
Db 150 LSRAYASNLGGYKNEGFEVLAAQOSPENPNWFGQTADAVRQYLWLFEEHNVLVLAG 209
QY 198 ---RFVQERQOLGRPVFPKAIKELHGTGIAIRLMLFYARAAAGIVDPAHGTGVANAGLVY 253
Db 198 ---RFVQERQOLGRPVFPKAIKELHGTGIAIRLMLFYARAAAGIVDPAHGTGVANAGLVY 253
QY 210 DILYRMDYER-----FIQAHRESADITVASLPMDERATA-----FGLMK 250
Db 210 DILYRMDYER-----FIQAHRESADITVASLPMDERATA-----FGLMK 250
QY 254 EN--GRLIAME-----DD-----LPVQVQITPNG-----D 277
Db 254 EN--GRLIAME-----DD-----LPVQVQITPNG-----D 277
QY 251 IDEGRIVESEKPKGQLKAMKVDITILGDDERAKEMFY---IASMGIVYVSKHVMLD 307
Db 251 IDEGRIVESEKPKGQLKAMKVDITILGDDERAKEMFY---IASMGIVYVSKHVMLD 307
QY 278 L---KTVGRFDFDQLESTMIAPKVDPSGEL---FALSVD----- 313
Db 278 L---KTVGRFDFDQLESTMIAPKVDPSGEL---FALSVD----- 313
QY 308 LLRDKFPFGANDFGSE-----VIPGATELGRVQAYLDGYWEDIGTIEAFYNANL 357
Db 308 LLRDKFPFGANDFGSE-----VIPGATELGRVQAYLDGYWEDIGTIEAFYNANL 357
QY 314 -VWSKP---YLKYFRFSPDGTSPDYELQDQTMHDFALTENFV---VFPDQV--- 362
Db 314 -VWSKP---YLKYFRFSPDGTSPDYELQDQTMHDFALTENFV---VFPDQV--- 362
QY 358 GITKPKVPDFSYDRSSPIYTPQ-----RYLPPSKMLDADITDSVIGEGCVIKNCKIHS 412
Db 358 GITKPKVPDFSYDRSSPIYTPQ-----RYLPPSKMLDADITDSVIGEGCVIKNCKIHS 412
QY 363 VFKLPEMI-----RGSPVYVYDKNKNVARFGLDKYAE 394
Db 363 VFKLPEMI-----RGSPVYVYDKNKNVARFGLDKYAE 394
QY 413 VVGLRSCISEGAIIEDLLMGADYETDADRRLAAGGVPIGIGKNSHKRAIDKNAR 472
Db 413 VVGLRSCISEGAIIEDLLMGADYETDADRRLAAGGVPIGIGKNSHKRAIDKNAR 472
QY 395 DSSNIKWIDAPDCFCFHLWNAKEPE-----TBEVVVIGSCMTPTPDSI 437
Db 395 DSSNIKWIDAPDCFCFHLWNAKEPE-----TBEVVVIGSCMTPTPDSI 437
QY 473 IGDVVKIINS-----NVQEAARETEGYFIKSGIVTVIKDALIPSGTV 515
Db 473 IGDVVKIINS-----NVQEAARETEGYFIKSGIVTVIKDALIPSGTV 515
RESULT 14
US-08-577-403-4
; Sequence 4, Application US/08577403
; Patent No. 5773693
; GENERAL INFORMATION:
; APPLICANT: Burgess, Diane G.
; APPLICANT: Dooner, Hugo K.
; TITLE OF INVENTION: Pea ADP-Glucose Pyrophosphorylase Subunit
; TITLE OF INVENTION: Genes and Their Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/577,403
; FILING DATE: 22-DEC-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: 08/097,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 12176-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-577-403-4

Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;

QY 1 MASFTATAVSGRWLGNHTQPPLLSSQSSDLSYCSLPLWASRYTKLVYSSA-----L 54
DB 1 MASMAIGVLK-----VPSSSSSSS-----SSSSSKAIARNLSFTSSOLGDKI 46

QY 55 HTPPALHFPKQSSNSPAIVVVKPKA-KESNTKOMNLFORAAAAALDAAGETLV---SHEKL 110
DB 47 FT---VSGIFRSGRNPFIVSPKAVSDSKNSQICLDPDASRVL-----GIILGGAGATRL 99

QY 111 HPL-PKTADPSVQIAGNFAVNEOPVRNPLPVVGKLPDSIKGYVVRNAGANPLHEPVTHH 169
DB 100 YPLTKRAKPAVPLGANRYLI-----DIPVSNCLNSISKIYVLTQFN---SASLNHR 149

QY 170 F-----PDGDMVHAKFE-----HGSA-----SVACRFTQIN----- 197
DB 150 LSRAYASNLGGYKNEGFEVFLAAQQSPENPNWFGTADAVRQYLWLFEHNVLLEYLVLAG 209

QY 198 ---RFVQERQLGRPVFPKAIAGELHGHGTGIARLMLFYARAAAGIVDPHAGTGVANAGLVY 253
DB 210 DHLRYMDYER-----FIQAHRESADITVASLPMDEARATA-----FGLMK 250

QY 254 FN--GRLLAMSE-----DB-----LPYQVOITPNG-----D 277
DB 251 IDEGRIVFESEKPKGEQLKAMKVDTITLGLDDEKAKEMPY---IASMGIVVYVSKHVMLD 307

QY 278 L---KTVGRFDFGQLESTMAHPKVDPESEL-----FALSXD----- 313
DB 308 LLRDKFPGANDFGE-----VTPGATELGLRVQAYLYGYWEDICTIEAFYNANL 357

QY 314 -VYVKP---YLKYFRFSPDGTSPDVEIQDQPTMMHDAITENFV---VVPDQCV--- 362
DB 358 GITKKVPDPFSFYDRSSPIVTPQ-----RYLPPSKMLDADITDSVIGEGCVIKNCXTHS 412

QY 363 VFKLPEMI-----RGSPVYVDKNKVARFQILDKYAE 394
DB 413 VVGLRSCISEGAIETLTMGADYYETDARRFLAAKGGVPIGKNSHIKRAIDKNAR 472

QY 395 DSSNIKWIDAPDFCFHLNWAEEPE-----TDEVVVVIGSCMTPTPDSI 437
DB 473 IGDVXIINSD-----NVQEAARETEGYFIKSGIVTVIKDALIPSGIV 515

RESULT 15

US-09-335-409-3
Sequence 3, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1410
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 3.0%; Score 96; DB 3; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.7;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;

QY 10 VSRWLGNGNHTQPPLLSSQS--SDLSY-----CSSLPWASKVTRK-----LNVSALHT 56
DB 645 VSEAGVEGDGDPMPMIQIOTPSDLAYIYTSGSTGLPKGVWIDHRGAVTILDINERFEI 704

QY 57 PPALHFPKQSSNS-----PAIVVVKPKA-KESNTKOMNLFORAAAA---AL 97
DB 705 GPGRVILALSSLSFDLSYDVVFGILAAAGTIVVDPASKLRDPAHWAELIEREKVTIVNSV 764

QY 98 DAAEGELVSHKELHP--LPKTADPSVQIAGNFAV---NEQPVRRNLPV---GKLPDS 148
DB 765 PALMRMLVEHFEGRPOSLSKLSLS--LSGDWIPVGLPGLQALRPGVSVISLGGATEAS 823

QY 149 I--KGIVVRNAGANPLHEPVTHGHFFDGDGMVHAKFEHSGSASYACRF--TOTNRFVQERQL 205
DB 824 IWSIGYFVRN-----VDLSWASIPYGRPLNQTFFHVLDEALE 860

QY 206 GRPVFKAIGELH-GHTGIARLMLFYAR-----AAAGIVDPAHGTGVANAGLYVFNGLL 259
DB 861 PRPVW--VPQOLYIGGVGLA---LGWWRDEEKTARKSLVHPETGERLYKTGDL---GRYL 912

QY 260 AMSEDDLPYQVOITPNGDLAKTVGREFD-----GQLESTMAHPKVDPESELFA 309
DB 913 -----PDGNIEMFGRDNQIKLRYRVSELGEIETETLSKHPNV---RDVIVP 956

QY 310 LSYDVVSKPYLYFRFSPDGTSPDVEIQDQPTMMHDEFA--ITENFVVVDPDQCVVFKLP 367
DB 957 VGDAANKLLLAY--VVPEGTRRRRAAQDASLKTTERIDARAHAAEADGLSDGERVQFKLA 1014

QY 368 EM-IR---GGSPVYVDKNKVARFQILDKYAE DSSNIKWIDAP 405
DB 1015 RHGLRRDLGKPVVDLTGQDPREAGLDVYARRRSVRTLEAP 1056

Search completed: May 19, 2003, 15:49:59

Job time : 45 secs

BEST AVAILABLE COPY

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:48:44 ; Search time 58 Seconds
(without alignments)
996.104 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGHT.....VPYGHGTFIGADDLAKQV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	US-09-758-269-6
2	2280.5	72.4	605	10	US-09-758-269-16
3	2168.5	68.8	612	10	US-09-758-269-12
4	1991	63.2	593	10	US-09-758-269-2
5	1930	61.3	604	10	US-09-758-269-14
6	1663.5	52.8	577	10	US-09-758-269-10
7	968	30.7	595	10	US-09-758-269-4
8	938	29.8	538	10	US-09-758-269-8
9	937	29.7	538	10	US-09-758-269-18
10	936	29.7	538	10	US-09-758-269-33
11	411.5	13.1	505	1	US-08-976-063C-22
12	260	8.3	526	9	US-10-053-192-1
13	247	7.8	506	9	US-10-053-192-4
14	185.5	5.9	529	9	US-10-053-192-5
15	107	3.4	26926	9	US-09-759-508B-2
16	106.5	3.4	2701	9	US-10-171-311-83
17	106	3.4	412	10	US-09-922-261-77
18	105	3.3	5701	10	US-09-864-761-37319
19	103.5	3.3	1311	9	US-10-103-377C-6

20	103	3.3	1156	10	US-09-815-242-13187	Sequence 13187, A
21	101.5	3.2	4342	10	US-09-815-242-5107	Sequence 5107, Ap
22	98.5	3.1	2828	9	US-10-176-847-54	Sequence 54, Appl
23	98.5	3.1	2828	10	US-09-905-129-21	Sequence 21, Appl
24	98.5	3.1	2828	10	US-09-991-630-21	Sequence 21, Appl
25	98	3.1	748	9	US-09-272-975-2	Sequence 2, Appl
26	98	3.1	753	9	US-09-272-975-58	Sequence 58, Appl
27	97.5	3.1	411	9	US-09-738-626-3692	Sequence 3692, Ap
28	97.5	3.1	1302	9	US-09-738-626-5616	Sequence 5616, Ap
29	96	3.0	666	10	US-09-815-242-10929	Sequence 10929, A
30	96	3.0	1410	9	US-10-014-717-3	Sequence 3, Appl
31	95.5	3.0	788	10	US-09-815-242-10562	Sequence 10562, A
32	95.5	3.0	1802	10	US-09-965-553-18	Sequence 18, Appl
33	94	3.0	1080	10	US-09-904-380-2	Sequence 2, Appl
34	94	3.0	1322	9	US-10-083-357-1332	Sequence 1332, A
35	93	3.0	566	10	US-09-391-340-2	Sequence 2, Appl
36	92.5	2.9	868	10	US-09-948-369-2	Sequence 2, Appl
37	92.5	2.9	868	10	US-09-948-369-2	Sequence 2, Appl
38	92.5	2.9	1057	9	US-10-217-700-2	Sequence 2, Appl
39	92.5	2.9	1057	10	US-09-918-909-29	Sequence 29, Appl
40	92.5	2.9	1131	10	US-09-801-368-72	Sequence 72, Appl
41	92	2.9	1167	9	US-10-147-026-8	Sequence 8, Appl
42	91.5	2.9	1081	9	US-10-217-700-4	Sequence 4, Appl
43	90.5	2.9	729	10	US-09-919-497-51	Sequence 51, Appl
44	90.5	2.9	916	10	US-09-745-763-174	Sequence 174, App
45	90.5	2.9	1038	10	US-09-908-500A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-758-269-6

; Sequence 6, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 599

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; US-09-758-269-6

Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 3.5e-280;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASFTATAVSGRWLGNGHTOPPLSSQSDLSYCSSLPMASRVTRKLNYSALHTPPAL 60

Db 1 MASFTATAVSGRWLGNGHTOPPLSSQSDLSYCSSLPMASRVTRKLNYSALHTPPAL 60

Qy 61 HFPQSSNPATVVKPKAKESNTKQMNLFQRAAAALDAAGFLVSHKHLPLPKTADPS 120

Db 61 HFPQSSNPATVVKPKAKESNTKQMNLFQRAAAALDAAGFLVSHKHLPLPKTADPS 120

Qy 121 VQIAGNAPVNEQVRNRLPVWGKLPDSIKGVYVYRNCANPLHEPVTGHFFDGDGMVHAV 180

Db 121 VQIAGNAPVNEQVRNRLPVWGKLPDSIKGVYVYRNCANPLHEPVTGHFFDGDGMVHAV 180

Qy 181 KFEHGSASYACRFOTNRFVQERQGRPVFPKAIQELGHGTGIAIARMLFYARAAAGVDP 240

Db 181 KPEHGSASYACRFTQTNRFRVQERQLGRPVPPKKAIGELHGHGTIARLMLFYARAAAGIVDP 240
QY 241 AHGTGVANAGLVFENGRLAMSEDDLPYQVQITPNGDLKTVGREFDQGOLES 292
Db 241 AHGTGVANAGLVFENGRLAMSEDDLPYQVQITPNGDLKTVGREFDQGOLES 297
QY 241 AHGTGVANAGLVFENGRLAMSEDDLPYQVQITPNGDLKTVGREFDQGOLES 352
Db 241 AHGTGVANAGLVFENGRLAMSEDDLPYQVQITPNGDLKTVGREFDQGOLES 357
QY 301 DPESGELFALSVDVSKPYLKYFRFSPDGTGKSPDVEIQDQPTMMHDFATE 472
Db 301 DPESGELFALSVDVSKPYLKYFRFSPDGTGKSPDVEIQDQPTMMHDFATE 477
QY 361 QVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEDSSNKKWIDAPDCFCFHLNANDEPE 532
Db 361 QVFKLPMTIRGSGSPVVDKKNKVARFGLDKYAEDSSNKKWIDAPDCFCFHLNANDEPE 537
QY 421 TDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLNKLTGSTRPITISNEDQVNLNLEAGM 590
Db 421 TDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLNKLTGSTRPITISNEDQVNLNLEAGM 599
QY 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLVGDNRVGGPFLPGEGGE 540
Db 481 VNRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLVGDNRVGGPFLPGEGGE 549
QY 541 EDESGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRVPYGFHGTGICADDLAKQV 599
Db 541 EDESGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRVPYGFHGTGICADDLAKQV 599

RESULT 2
US-09-758-269-16
; Sequence 16, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-758-269-16
Query Match 72.4%; Score 2280.5; DB 10; Length 605;
Best Local Similarity 71.7%; Pred. No. 2.1e-200;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;
QY 5 TATAAAGRWLGNHQPPLSSSQSDLSYCS---SLPMASRVTRKLNLSVSSALHTPEPALH 61
Db 3 TTSHAINTWI-----KTLSPSSKEFGFASNSISLLKNQHRQSLNINSSLAQAPILH 57
QY 62 FPKQSSN--SPA--IVVKKPAKESN-----TKOMNLFQRAAAALDAARGLVSHKELHP 112
Db 58 FPKQSSNYQTPKNTTISHPKQENNNSSSTSKNVLQRAAAMALDAVESALTKHELEHP 117
QY 113 LPKTPADSPVQIAGNFAPVNEQPVRRNLPPVVGKLPDSITGVYVNGANPLHPEVPIGHFFD 172
Db 118 LPKTPADSPVQIAGNFAPVNEQPVRRNLPPVVGKLPDSITGVYVNGANPLHPEVPIGHFFD 177
QY 173 GDGMVHAFKFEHGSASYACRFTQTNRFRVQERQLGRPVPPKKAIGELHGHGTIARLMLFYAR 232
Db 178 GDGMVHAFKFEHGSASYACRFTQTNRFRVQERQLGRPVPPKKAIGELHGHGTIARLMLFYAR 237

QY 233 AAAGIVDPAHGTGVANAGLVFENGRLAMSEDDLPYQVQITPNGDLKTVGREFDQGOLES 292
Db 238 GLFGLVDHSGTGVANAGLVFENGRLAMSEDDLPYHVKVTPDGLKTEGREFDQGOLES 297
QY 293 TMIAHPKLPDVPSELGELFALSVDVSKPYLKYFRFSPDGTGKSPDVEIQDQPTMMHDFATE 352
Db 298 TMIAHPKLPDVPSELGELFALSVDVSKPYLKYFRFSPDGTGKSPDVEIQDQPTMMHDFATE 357
QY 353 NFVVVDQVQVFKLPEMIRGSGSPVVDKKNKVARFGLDKYAEDSSNKKWIDAPDCFCFHL 412
Db 358 NFVVVDQVQVFKLPEMIRGSGSPVVDKKNKVARFGLDKYAEDSSNKKWIDAPDCFCFHL 417
QY 413 WNAWEPEPTDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLNKLTGSTRPITISNEDQ 472
Db 418 WNAWEPEPTDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLNKLTGSTRPITISNEDQ 477
QY 473 QVNLKAGMVRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLVGDNRVGGPFL 532
Db 478 QVNLKAGMVRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLVGDNRVGGPFL 537
QY 533 FLPGGE--GGEDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRVPYGFHGTG 590
Db 538 FLPGGE--GGEDEGYILCFVHDEKTKWSELOIYNVNSLEVEATVKLPSPRVPYGFHGTG 597
QY 591 ADDLAKQ 597
Db 598 ANDLANQ 604

RESULT 3
US-09-758-269-12
; Sequence 12, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Vigna unguiculata
US-09-758-269-12

Query Match 68.8%; Score 2168.5; DB 10; Length 612;
Best Local Similarity 68.6%; Pred. No. 4e-190;
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;
QY 8 AAVSGRWLGNHQPPLSSSQSDLSYCS---SLPMASRVTRKLNLSVSSALHTPEPALH 65
Db 3 SSASNTWTFNATLSPSPFPKLPSTISPTNLLPL--RKTSSNTITCSLQT---LHFPKQY 56
QY 66 ----SSNSFAIVVKKPAKESNT-----KOMNLFQRAAAALDAARGLVSHKELHP 112
Db 57 QPTSTSTATTATTTTPPIKTTTITTTTPPTNPLSDTNPQLPKQKNFLQKAAATDLV 116
QY 101 EGFVLSHEKHLPLPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSITGVYVNGANPL 160
Db 117 ETALVSHKHLPLPKTADPSVQIAGNFAPVNEQPVRRNLPPVVGKLPDSITGVYVNGANPL 176
QY 161 LHPVPTVGHFFDGDGMVHAFKFEHGSASYACRFTQTNRFRVQERQLGRPVPPKKAIGELHGH 220
Db 177 LHPVAGHFFDGDGMVHAFKFEHGSASYACRFTQTNRFRVQERQLGRPVPPKKAIGELHGH 236

Db 252 LVTFNRLAMSEDDLPYHVRVADGDLTEVGRYDGGQGCCAMIAHAKLPATGELHAL 311
QY 311 SYDWSKPYLKYFRSPDGTOKSPDVEIQDPTMMHDFALTENFVVVVDQOVVFKLPPEMI 370
Db 312 SYDVIRKPYLYEYFRPDGTSKDDVEIPLDPTMIHDFALTENFVVVVDQOVVFKLQEML 371
QY 371 RGSPPVVDKKNVARGILDKYAEDSSNIKWIDAPDCFCFHLNANWEPEPTDEVVIGSC 430
Db 372 RGSPPVVDKREKTSRGLVLPKHAADASEMAWVDVDPDCFCFHLNANWEDEATGEVVVIGSC 431
QY 431 MTPDPSIFNESDENLKSLSVSEIRLNKLTGSTRPPIISNDQOVNLEAGVNRNMLGRKI 490
Db 432 MTPADSFIFNESDELESVLTIRLDARTGRSTRVLP-PSQENLEVGVMNRNLLGRES 490
QY 491 KFYALALAEPPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 545
Db 491 RYALAVAEPPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 550
QY 546 ILCFVHDEKTKSELOTVNAVSLVEATVKLPSPVPGFHTGTFIGADDLAKQ 597
Db 551 VLFVHDERAGTSELLVYVNAADIRLEATVQLPSRVPFGFHTGTFITGQELEAQ 602

RESULT 6

US-09-758-269-10

; Sequence 10, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 10

; LENGTH: 577

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-10

Query Match 52.8%; Score 1663.5; DB 10; Length 577;
Best Local Similarity 55.4%; Pred. No. 7.8e-144;
Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

QY 24 LSSQSSDLSYCSLIPMASRVTRKLVNVS-ALHTPPALHPFKQSSNSPAIVVVKPKAKESN 82
Db 9 LPIPTSPRSHLLPOPKNANISRRLLINPFKIPLDITSPVSP-----VKLQPTYPN 62
QY 83 TKOMLFORAAAAALDAE-GFLYSHEKHLPLKPTADPSVOIAGNFAPVNPQVRRNLPV 141
Db 63 ---LNLQKLAATMLDKTIESSIVIPMEQNRPLPKPTDPAVOLSGNFAPVNECPVQNGLEV 119
QY 142 VGLKPDSTKGYVVRNAGNPLHEPVYTGHHFFDGDGMVHAVK--FEHGSASYACRFTQTNR 199
Db 120 VQGPISCLKGVYIRNAGNPMFPPLAGHHLFDGDGMHIAVSTGFON-QVSYSCRVTKTNR 178
QY 200 VQEROLGRVPFKALGELHGHGTGIALMLFYARAAGIVDPAHGTGVNAGLVYFNGRL 259
Db 179 VQETALGRVFPKPIGELHGHSLARLALFTARAGIGLVGTRGMVANAGVFFNGRL 238
QY 260 AMSEDDLRYQVQITPBGDKLVGREFDQGLQESTMIAHPKVDPSGELFALSVDVSKPY 319
Db 239 AMSEDDLRYQVQITPBGDKLVGREFDQGLQESTMIAHPKVDPSGELFALSVDVSKPY 319

QY 320 LKYREFSPDGTOKSPDVEIQDPTMMHDFALTENFVVVVDQOVVFKLPPEMIRGSPVVD 379
Db 299 LRYLKTCGKTRDVEITLPEPTMIHDFALTENFVVVVDQOVVFKLSEMRGSPVIYV 358
QY 380 KKNVARGILDKYAEDSSNIKWIDAPDCFCFHLNANWEPEPTDE----VVVIGSCTMPPD 435
Db 359 KEMARFGLSKQDLTGSIDINWVDVDPDCFCFHLNANWE-RTGEGDPVIVVIGSCMSPD 417
QY 436 STINESDENLKSLSVSEIRLNKLTGSTRPPIISNDQOVNLEAGVNRNMLGRKTFAYL 495
Db 418 TIFSEGEPTFVSEIRLNKLTGSTRPPIISNDQOVNLEAGVNRNMLGRKTFAYL 473
QY 496 ALAEPMPKVSFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGEDEGYILCFVHDEKT 555
Db 474 AIADPPKPCSGIAKVDIQNTGSEFNYGSRFGGEPCEVPEGEDEKGYVMGFVRDEK 533
QY 556 WKSELQTVNAVSLVEATVKLPSPVPGFHTGTFIGADDLAKQV 598
Db 534 DESEFVVVDATDMKQVAAVRLPERVPYGFHTGTFVSNQLKEQV 576

RESULT 7

US-09-758-269-4

; Sequence 4, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; PRIOR FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 4

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-758-269-4

Query Match 30.7%; Score 968; DB 10; Length 595;
Best Local Similarity 36.7%; Pred. No. 5.3e-80;
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVNVSALHTPPALHPFKQSSNSPAIVVVKPKAKESNTKOMLFORAAAAALDAE 101
Db 42 SPITNPSSNNRRNKKPKTLH---NRTNHTLVSSPPKLRPEMTALALF---TTVEDVIN 94
QY 102 GFLYSHEKHLPLKPTADPSVOIAGNFAPVNPQVRRNLPV-GKLPDSIKGYVVRNAGN 160
Db 95 TFDIP-----FSRSVDPKHVLSONFAPVLDLPDPTDCEIHTGLPLSLNAYIRNGENP 149
QY 161 LHPEVTGHHFFDGDGMVHAVKFEHGSASYACRFTQTNRFPVQEROLGRVPFPAIGELHGH 220
Db 150 QFLPRGPHVLEFDGDGMHIAKIHNGKATLCSRYKVTYKYNVEKOTGAPVMNPFSGENG 209
QY 221 T-GIARLMLFYARAAGIVDPAHGTGVNAGLVYFNGLLAMSEDDLPYQVQITPBGDLK 279
Db 210 TASVARGALTAARVLITGOYNPVGNGIGLANTSLAFFSNLFFALGESDLPYAVRLTESG 269
QY 280 TVGREDDQGLQESTMIAHPKVDPSGELFALSVDVSKPYLKYREFSPDGTOKSPDVEI-Q 338
Db 270 TIGRTDFDQGLKAMSTAHPKTDPTITGETFAFYGPV-PPFLTYFEDSAGKKQDVPIFS 328
QY 339 LDQPTMHDFALTENFVVVVDQOVVFK-----LPMEIRGSPVYVYDKNKVARFGLDKYAE 395
Db 329 MTPSPFLHDFALTENFVVVVDQOVVFK-----LPMEIRGSPVYVYDKNKVARFGLDKYAE 388

Db 344 EDEVVLIITCRLENPLDMDVSGKVKLENGNELYEMRNMKTGSASOKKLASASA----- 398
Qy 476 LEAGMVNRMGLGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDFPRINCYTGKKORYVYGTILDSIAKVIKILFDLHAEATGKRMLEVGNGIKGIYDL 458
Qy 523 GNRVGGEPFLFPGGGREDEGYILCFVHDEKTKWSELIQIVNAVSLVB--ATVKLPSRV 580
Db 459 GEGRGSEALYVPRETAEDDGYLFFVHDENTGKSCVTVIDAKTMSAEPVAVVLPHRV 518
Qy 581 PYGFHGTFTGADDLAKQVV 599
Db 519 PYGPHALFVTEEQLOEQTL 537
RESULT 9
US-09-758-269-18
; Sequence 18, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-18
Query Match 29.7%; Score 937; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 3,1e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
Qy 64 KOSSNSPAIVVKKPK-AKESNTKQMNLFQRAAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
Db 4 KLSGSSIIIVHPRESKGFSSKLLDLRLRVVKLM-----HDASLPLH-----Y 47
Qy 123 IAGNFAPV-NEQPVRRNLPPVYKLPDSIKGVYVRNGANPLHEPVYTGHHFFDGDGMVHAYK 181
Db 48 LSGNFAPRDETPPVYKLPDSIKGVYVRNGANPLHEPVYTGHHFFDGDGMVHAYK 107
Qy 182 FEHGSASYACRFTQNRVQERQLGRPVFPKAIAGELHGTGIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATVSRVYKTSRLKQEEFFCAAKFMK-IGDLKFGFGLLMVYQOOLRKLKLDNT 166
Qy 242 HGTGVANAGLVYNGRLLAMSEDDLPYQVQITPNDGLTKVGRFDGQLESTMIAPKVD 301
Db 167 YNGTANTALVYHGGKLLALQADRPYVTKVLEDGLQTLIDYDKRLTHSFTAHKVD 226
Qy 302 PESGELFALSYDVSKPYLYKRFSPDGTSPDVELQDQPTMMHDFATENFVVVPDQO 361
Db 227 PVTFEMFTGYS-HTPPYLYTVRISKDGIHDPVPVITISEPTMMHDFATETAYFMDLP 285
Qy 362 VYFKLPEMIRGSGSPV-YDNKVARFGILDYAEDESSNKKWIDAPDCFCFHLNAREEPE 420
Db 286 MHFRPKEMVKKMIYSFDPKAREGVLPYAKOELMIRWELPNCFLFHNANAW--E 343
Qy 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSIVLSIRLNKLTGSTRPIISNEDQOYN 475
Db 344 EDEVVLIITCRLENPLDMDVSGKVKLENGNELYEMRNMKTGSASOKKLASASA----- 398
Qy 476 LEAGMVNRMGLGRKTKFAYLALAEWPVKVSGFAKVD-----LTTGEVKKHLY-- 522

Qy 396 SSNIKWIDAPDCFCFHLNAREEPEDEVVVIGSCMTPPDSIFNESDENLKSIVLSIRLN 455
Db 399 ESEKWFEPVGNIIHA,NAWEDDNGSVVLIAPNIMSIEHTLERMD-LVHALVEKVKID 447
Qy 456 IATGSTRPIISNEDQOYNLEAGMVNRMGLGRKTKFAYLALAEWPVKVSGFAKVDLITG 515
Db 448 LVTGIVRRHPISAR-----NIDFAYINAFPLGRCSRYVYAAIGDPMPKISGVVVKLDSK 502
Qy 516 E-----VKKHLGDNRYGGEPLFLPEGG-----EDEGYILCFVHDEKTKWSELIQIVNAV 566
Db 503 DRDDCTVARMYGGSCYGGEPFVARDPQNPABEDDGYVVIYVHDEVTEGESKELYMDAK 562
Qy 567 S--LEVEATVKLPSRVYGFHGTFTGADDLAK 596
Db 563 SPELEIVAARLPRPVYGFHGLFVKESDLNK 594
RESULT 8
US-09-758-269-8
; Sequence 8, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-8
Query Match 29.8%; Score 938; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 2.5e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
Qy 64 KOSSNSPAIVVKKPK-AKESNTKQMNLFQRAAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
Db 4 KLSGSSIIIVHPRESKGFSSKLLDLRLRVVKLM-----HDASLPLH-----Y 47
Qy 123 IAGNFAPV-NEQPVRRNLPPVYKLPDSIKGVYVRNGANPLHEPVYTGHHFFDGDGMVHAYK 181
Db 48 LSGNFAPRDETPPVYKLPDSIKGVYVRNGANPLHEPVYTGHHFFDGDGMVHAYK 107
Qy 182 FEHGSASYACRFTQNRVQERQLGRPVFPKAIAGELHGTGIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATVSRVYKTSRLKQEEFFCAAKFMK-IGDLKFGFGLLMVYQOOLRKLKLDNT 166
Qy 242 HGTGVANAGLVYNGRLLAMSEDDLPYQVQITPNDGLTKVGRFDGQLESTMIAPKVD 301
Db 167 YNGTANTALVYHGGKLLALQADRPYVTKVLEDGLQTLIDYDKRLTHSFTAHKVD 226
Qy 302 PESGELFALSYDVSKPYLYKRFSPDGTSPDVELQDQPTMMHDFATENFVVVPDQO 361
Db 227 PVTFEMFTGYS-HTPPYLYTVRISKDGIHDPVPVITISEPTMMHDFATETAYFMDLP 285
Qy 362 VYFKLPEMIRGSGSPV-YDNKVARFGILDYAEDESSNKKWIDAPDCFCFHLNAREEPE 420
Db 286 MHFRPKEMVKKMIYSFDPKAREGVLPYAKOELMIRWELPNCFLFHNANAW--E 343
Qy 421 TDEVVVIGSCMTPPD-----SIFNESDENLKSIVLSIRLNKLTGSTRPIISNEDQOYN 475

Db 399 VDFPRINECYTGKQRYVYVGTILDSIAKVTGIKFDLHAFATGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWSELOIVNAVSLEVE--ATVKLPSPRV 580
Db 459 GGRYGSEAIYVPRETAEDDGYLIFVHDENIGKSCVTVIDAKTMSAPVAVVELPHRV 518
QY 581 PYGFHGTGFIAGDADLAKOVV 599
Db 519 PYGFHALFVTEOLOEQTL 537
RESULT 10
US-09-758-269-33
; Sequence 33, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 33
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-33

Query Match 29.7%; Score 936; DB 10; Length 538;
Best Local Similarity 37.0%; Pred. No. 3,9e-77;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;
QY 64 KQSSNPAIVVVK-AKESNTKOMNIFORAAAALDAAGFLVSHKHLPLKTAQPSVQ 122
Db 4 KLDGSIITISVHPRKSGFSLDLRLVVKLM-----HDASLPLH-----Y 47
QY 123 IAGNFAPV-NEQPVRRNLPLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAVK 181
Db 48 LSGNEAIPRDETPVKDLPVHGFLPECLNGEFVRVGNPKFDVAGVHWFDDGDMHIGVR 107
QY 182 FEHGSYACRFTQTNRFVQERQLGRPVFPKAIAGELHGHGTGIARLMLFYARAAAGIVDPA 241
Db 108 IKDGKATYVSRVYKTSKLOEPEFGAAKPMK-IGDLKGGFFGLLMVNYQQLFKLILNT 166
QY 242 HGTGVNAGLVFNGRLAMSEDDLPVOVQITPNDLKTVGRFEDGQLESTMIAPKVD 301
Db 167 YGNGTANTALVYHGLLALQEAQKPVVIXVLEDGLQTLGIIDYDKRLHSTAHKVD 226
QY 302 PESGEIFALSYDVSVKPYLYKFRFSPDGTSPDVEIQDQTMHDPFAITENFVVPDQ 361
Db 227 PVIGEMFTFGYS-HTPPLYIRVLSKDGIMHDPVPIITSEPIIMHDPFAITETVAIENDLP 285
QY 362 VPKLPDMIRGSPV-VYDNKVARFGLDKYAEEDSSNIKWIDAPDCFCFLHNAWEEPE 420
Db 286 MHPRPKEMVKKMIYFDPDKARFGLPRYAKDELMIRWELPNCFFIFHNANAWP--E 343
QY 421 TDVVVIGSCMTDPD-----SIFNESDENLKSVLSEIRLNKLTGSTRRPIISNDQVN 475
Db 344 EDEVVLTICLENDDLMVSGKVKAEKLENFCNELYEMRNFNMKTSASQKLSASA----- 398
QY 476 LEAGMVNRNMLGRKTKFAYLALAPWPVKSGFAKVD-----LTTGEVKKHLY-- 522
Db 399 VDFPRINECYTGKQRYVYVGTILDSIAKVTGIKFDLHAFATGKRMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWSELOIVNAVSLEVE--ATVKLPSPRV 580
Db 459 GGRYGSEAIYVPRETAEDDGYLIFVHDENIGKSCVTVIDAKTMSAPVAVVELPHRV 518
QY 581 PYGFHGTGFIAGDADLAKOVV 599
Db 519 PYGFHALFVTEOLOEQTL 537
RESULT 11
US-08-976-063C-22
; Sequence 22, Application US/08976063C
; Publication No. US20020182697A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLI
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTRA
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,063C
; FILING DATE: 21-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 196 49 655.1 (Germany)
; FILING DATE: 29-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-976-063C-22

Query Match 13.1%; Score 411.5; DB 1; Length 505;
Best Local Similarity 25.8%; Pred. No. 4.6e-29;
Matches 137; Conservative 79; Mismatches 192; Indels 123; Gaps 19;
QY 118 DPSVQIAGNEFAPVNEQPVRRNLPLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMV 177
Db 8 DP--QLVGTLLPRTIADLDFDLEVDGEIPKISINGTFYRNTPEPQVTPQKEHTFIDGDMGMA 65
QY 178 HAVKFEHGSYACRFTQTNRFVQERQLGRPVFPKAIAGELHGHGTGIARLMLFYARAAAGI 237
Db 66 SAFHFEDGHVDFISRWVKTARFATRLARKLSLFGMYRNPYTDOTSVKGL----- 114
QY 238 VDBAHTGTVNAGLVFNGRLAMSEDDLPVOVQITPNDLKTVGRFEDGQLES-TMIA 296
Db 115 -----DRTVANIISIHHGKVLAVKEDGLPY--ELDPR-TLETGRGHFDYDQVTSQHTA 166
QY 297 HPKVDPESGSEFALSYDVWSKPYLVKFRFSPDGTSPDV-----ETQDQD- 342
Db 167 HPKYDPTGTGLL-----PFCSAKGEATPDMAIYIVDKHKVTHTWPEQPY 213

[illegible]

```

QY 395 DSSNIK-WIDAPDCFCFLHNAWEPEPTEVVVIGSCMTPPDSIFN-----ESD 442
Db 292 KEVSTKFTYDA--LVLYHHINAYEE--DGHVVFIVAYRDNLSYDMFTYKLLDKDFEYN 346
QY 443 ENLKSVL-----SEIBNLN-----KGTSETRRPLISNQDVNL 476
Db 347 NKLRSIPTCKRFVVPLOYDKDAEVSGLVKLPTSATAVKEKDGSIYQPEILCE----GI 402
QY 477 BAGVNNRMLGRKTKFAYLALAEWPKYSGFAKVLDLTITGEVKKHLYGDNRYGEGPEFLPG 536
Db 403 ELPRVNDYNGKKYKVVATEVQWSPVTKIAKLNQVTKEV--LHWGDEHCPSEIFVPS 461
QY 537 -EGGEDEGYIL-CFVHDEKTIWKSQLOTVNAVSLEV---EATVKLPSRPYPGPHCTIGAD 592
Db 462 PDAREDEGGVLTCTVVVSEPNKAPELLILDATKFKELGRAIVNVEHML--DLHGMEIFQN 519
QY 593 DLAKQ 597
Db 520 DLGAE 524

RESULT 13
US-10-053-192-4
: Sequence 4, Application US/10053192
: Publication No. US20030087336A1
: GENERAL INFORMATION:
: APPLICANT: BACHMANN, Heinrich
: APPLICANT: BRUGGER, Roland M
: APPLICANT: FRIEDLEIN, Arno M
: APPLICANT: WIRTZ, Gabriele M
: APPLICANT: WOGGON, Wolf-Dietrich
: APPLICANT: WYSS, Adrian
: APPLICANT: WYSS, Markus
: TITLE OF INVENTION: BETA,BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
: TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
: FILE REFERENCE: B.B-CAROTENE 15,15'-DIOXYGENASES,...
: CURRENT APPLICATION NUMBER: US/10/053,192
: CURRENT FILING DATE: 2002-01-15
: PRIOR APPLICATION NUMBER: 103382.0
: PRIOR FILING DATE: 1999-02-22
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 506
: TYPE: PRT
: ORGANISM: CHICKEN
US-10-053-192-4

Query Match 7.8%; Score 247; DB 9; Length 506;
Best Local Similarity 24.3%; Pred. No.5.7e-14;
Matches 130; Conservative 83; Mismatches 213; Indels 110; Gaps 29;

QY 132 EQPVRRNLFVVVGKLPDSITKGVVVRNGANPLHE--EVTGHHFFDGGMVHAVFEGHSASY 189
Db 2 EHPEPIKAEVQGLPTWLQGVLLRNGPG-MHTIGDTKYNHWFEDGLALLHSFTFKNGEVY 60
QY 190 ACRP-----TQTNRFVQERQLRPVFPKAIGELHGHGTGIARIMLFYARAAGIVDP 240
Db 61 RSKYLRSDTYNCNTEANRIVVS-EFGTWAYPDPCNNI-----FAKAFSYLSHTIPEFTDN 114
QY 241 AHGTVANAGLVFYNGRIAMSEDDLPYQVOITPNGLTKTYGREDFDQLESTM-IAHPK 299
Db 115 CL-INIMKTGDYY-----ATSEIN--FIRKIDQP-TLETLDKVDSKYVAVNATSPH 165
QY 300 VDPSEGFALSVDVWSPKYLKFRFS-----PDGKSPD----VEI-----OLDQPT 343
Db 166 YD-SAGNLTNMGTSIVDXGRGKYVLFKLPSSVPEKEKKKSCFKHLEVCVSPSRSLQPS 224
QY 344 MMHDEAITEENVVVDQOVFKL-----PEWIRG---GSPVVDKNKVARFGILDKYAE 394
Db 225 YYHSFGITENIVFLEQP--FKLDIVTKIATAYIRGVNNAWASCLSFKEKDKTWHFHYDRKTK 282
QY 395 DSSNIK-WIDAPDCFCFLHNAWEPEPTEVVVIGSCMTPPDSIFN-----ESD 442

```

Db 283 KEVSTKTYTDA--LVLXHNAYEE--DGHVDFDIAYVRDNSLYDMFYLLKLLDKDFEVN 337
QY 443 ENLKSVL-----SETRLN-----KTGSTRPPIISNQOQVNL 476
Db 338 NKTSITCKRFVVPLOYDKDAEYLSNLVKLPSTATAVKEKDGSIYQPELICE---GI 393
QY 477 EAGNVNMLGRKTKFAYLALAEWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPFLPLPG 536
Db 394 ELPRVNDYNGKKYKYAYATEVQVSPPTKAKLNQVKEV-LHWGEDHCWSPPIFVPS 452
QY 537 -EGGEDEGYIL-CFVHDEKTKWSELOIVNAVSLV-EATVKLPSPVPGFHGTF 588
Db 453 PDAREDEGVVLTCCVVVSEPNKAPFLLLDAKTEKELGRATVNVEMHL--DLHGWF 506

RESULT 14

US-10-053-192-5
; Sequence 5, Application US/10053192
; Publication No. US20030087336A1
; GENERAL INFORMATION:
; APPLICANT: BACHMANN, Heinrich
; APPLICANT: BRUGGER, Roland
; APPLICANT: FRIEDLEIN, Arno M
; APPLICANT: WIRTZ, Gabriele M
; APPLICANT: WOGGON, Wolf-Dietrich
; APPLICANT: WYSS, Adrian
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: BETA, BETA-CAROTENE 15, 15'-DIOXYGENASES, NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE
; FILE REFERENCE: B, B-CAROTENE 15, 15'-DIOXYGENASES, ...
; CURRENT APPLICATION NUMBER: US/10/053, 192
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 103382.0
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 529
; TYPE: PRT
; ORGANISM: BOVINE
US-10-053-192-5

Query Match 5.9%; Score 185.5; DB 9; Length 529;

Best Local Similarity 20.2%; Pred. No. 2.7e-08;
Matches 112; Conservative 91; Mismatches 209; Indels 143; Gaps 26;

QY 141 VVGKLPDSIKGVYVRNGANP-----LHEPVTGHHFFDGDGMVHAKPEHGSASVACRFTQ 195
Db 11 VTGRIPLWLGSLLRCTGCLPEVSGEPF--YILFDQALLHKFDEKGGHVTYHPEIR 68
QY 196 TNRFVQERQLGRPVFPAKIGELHCHTGIARMLPYARAAAGIVDPAHGTGVANAGLYFN 255
Db 69 TDAYVRAMTEKRIIVITE-----FGFTTCAPDPCC--KNIFSRFFSYER 109
QY 256 G-----RLAMSED-----DLPOVQVLTNGDLKTVGRDFDGOLE-----STM 294
Db 110 GVEVTDNALNVIPVGGEDYACTETNITKINPE-TLEII-----FTKQVLDCLNTVSNGA 164
QY 295 IAHPKVDPPES-----GELFALSVDVYSKPKYLFRESPOGKTSPQVVEIQLD----- 340
Db 165 TAPHIENDGTVYNIGNCFGNFSIAINVKIPLOADKEDPISKFSIENVVQPCSDRF 224
QY 341 OPTMMHDFATENVVVDQOVFKLEPMIRGGSPVYVKNKVARF-----GI 388
Db 225 KPSVHSGFLTPNVIIVETPVKINLEKFLSSWS--LWGANVMDCFESFNETMGVWLHI 282
QY 389 LDKVAEDSSNLKWDAPDCFCFHLNNAWEPETDEVVVICSMTPPDSINESD---ENL 445
Db 283 ADKARKKYLANKYTSP-FNLFHINTYE--DNGFLIVDLCCKWKGFEVINYFTLYLANL 339
QY 446 KSVLSEIRLNKLTGSTRRPTISNQOQVNLNAGMWRNM--LGRKTKFAYLA-----496

Db 340 REMEEVK---KNARKAPOPEVRVRYVPLNIDKADTKGNLVTLPNTATATILCSDEFTTI 396
QY 497 LAEPWPKVSG-----FAKVDLT--TGEVKKHLYG----- 523
Db 397 WLEPEVLSGPRQAFEPFOINQYCGKPYTYAYGLGLNHFPDRCLCKLNKTKETWFTV 456
QY 524 --DNRYGGEPFLFP-EGGEDEDEGYILCFVHDEKTKW--SELOIVNAVSL-----EYD 573
Db 457 WOEPDSTPSEPIFVSHPDALDEDDGVVLSVVVSPGAGQKPAYILLILNAKLSEVARAEFT 516
QY 574 VKLESRYVPYGFHGT 588
Db 517 VEI--NIPVTFHGLF 529

RESULT 15

US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759, 508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 3.4%; Score 107; DB 9; Length 26926;

Best Local Similarity 19.0%; Pred. No. 2.2e+02;
Matches 131; Conservative 97; Mismatches 205; Indels 256; Gaps 38;

QY 51 SSALHTPPAL-HPPKQSSNSPAI-VVKPKAKESNTKOMNL-----FORAAAAALDAAE 103
Db 15827 STYLNSEPTVAQYFVKVPGPGTPTVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERN 15886
QY 104 LVSHEKLIH--PLPKTADPSVQIAGNFAVNAEQPVV-----RNLVVGKLPDSIKGVY 154
Db 15887 SILWKLNKTPIPOTKTKTGL-----EGVEYEFVRSAAENIVGIGK-PSKVSECV 15937
QY 155 -RNGANPLHEPVTGHHFFDGDGMV---HAVKFEHGSASVACRFTQTNRFVQERQLGRPV 210
Db 15938 ARDPCDPPGRP-----EAIIVTRNSVTLOWKKTYDGGSKITGYIVKEKKELPEGRW 15988
QY 211 PKA-----IGELHGHGTGIARMLFYARAAAGIV-DPAHGTGVANAGLVYFNGR 257
Db 15989 MKASFNIIDHFEVTVGLVEDHRYEFVFI---ARNAGVSESESTGA----- 16034
QY 258 LLAMSEDDLPYQVOITPNGDLKTVGRDFDGOLESTMIHAKPKVDPESGELFALSVDVSK 317
Db 16035 IARDEVDPP-----RISMDPKYKDTIVVH-----AGESFKVDADIYK 16073
QY 318 PY--LKVFRSPDGTKSPDVEIQ-----LQOPTMMHDFATEN 353
Db 16074 PIPTQWIKGQDELNSFARLEIKSTDFATSLSVKDAVRVDSGNVYLKAKKAVAGERSVTN 16133
QY 354 FVVV-----PDQOVVEK-----LPEMIRGGSPVYVKNKVARFGLIOKYAEDSS 397
Db 16134 KVLDLRDPGPEGPVVISVTAECTLAWKPLQDGGSDII--NYIVE-----RRETS 16183
QY 398 NIKW--IDAP----DCFCFHLNNAW-----EPETDEVVVICSMTPPDS- 436
Db 16184 RLVTVDANVQTLSCCKVTKLLEGNETFRIMAVNKYGVGEPLSEPVVAKNPFVDPAD 16243
QY 437 -----IFNESDEN-----LKSVLSEIRLNKLT 458

Db 16244 KAPEVTTVTKDSMIWVWRPASDGGSEILGYVLEKRDKEGIRWTRCHKRLIGELRLRV-T 16302
Qy 459 GESTRPIISNEDQOVNLEAGMVNRMGLRKTKEAYLALAEFPWKVSGFAK----- 509
Db 16303 G-----LIENHDYEFVSA-----ENAAAG-----LSEFPSP-SAYQACDEIYKPG 16342
Qy 510 -----VDLTGTGEV-----KKHLYGDNRYGGEPFLPGEGGEDEGYIL--CFVH-DEKT 555
Db 16343 PPNNPKVIDITRSSVFLSWSKPIY-----DGGCEIOGYIYKCDVNWGEMT 16388
Qy 556 WKSELQIVNAYSLEVEATVKLPSRVPYGF 584
Db 16389 MCTPPTGINKTNIEVE---KLEKHEYNE 16414

Search completed: May 19, 2003, 15:58:18
Job time : 77 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2289	72.7		604	2	T51936	probable 9-cis-epo
2	2280.5	72.4		605	2	T07123	nine-cis-epoxy caro
3	2170.5	68.9		657	2	E96812	protein t3f9.10 [i
4	2044.5	64.9		589	2	A86425	probable 9-cis-epo
5	1991	63.2		583	2	T04531	nine-cis-epoxy caro
6	1930	61.3		604	2	T04351	viriparous-14 prot
7	968	30.7		595	2	T04438	hypothetical prote
8	938	29.8		538	2	T04939	neoxanthin cleavag
9	937	29.7		538	2	T51734	neoxanthin cleavag
10	745.5	23.7		475	2	AG1944	hypothetical prote
11	745	23.7		446	2	T17019	hypothetical prote
12	528.5	16.8		483	2	E87345	conserved hypothe
13	466	14.8		501	2	A70534	hypothetical prote
14	454.5	14.4		502	2	A70582	hypothetical prote
15	394	12.5		490	2	S76169	hypothetical prote
16	355.5	11.3		497	2	AE2341	hypothetical prote
17	352	11.2		484	2	JN0595	lignostilbene alph
18	337	10.1		480	2	S76206	lignostilbene alph
19	303.5	9.6		489	2	JC4324	hypothetical prote
20	253	8.0		616	2	T10688	lignostilbene alph
21	244	7.7		618	2	CA3485	hypothetical prote
22	242	7.7		472	2	AG2417	hypothetical prote
23	238.5	7.6		556	2	F88115	hypothetical prote
24	231	7.3		483	2	D87290	protein F53C3.12 [
25	215	6.8		533	2	A47143	conserved hypothe
26	126	4.0		790	2	T34293	retinal pigment m
27	116.5	3.7		1932	2	T31113	hypothetical prote
28	114	3.6		870	2	S74291	mucin-like glycopr
29	113	3.6		466	1	IQE9V	hypothetical prote
30	113	3.6		466	1	IQE9V	replication initia

Db 418 NAWPEPTDEIVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 477
QY 474 VNLEAGMVRNMLGRKTKFAYLALAEAPKPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPLF 533
Db 478 VNLEAGMVRNMLGRKTKFAYLALAEAPKPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPLF 537
QY 534 LPGE--GGEDEGYILCFVHDEKTKWSELOITVNAVSELEATVKLPSPVPGPHGTGIGA 591
Db 538 LRPDPNSKEEDGYILAFVHDEKTKWSELOITVNAVSELEATVKLPSPVPGPHGTGIGA 597
QY 592 DDLAKQ 597
Db 598 NDLANQ 603

RESULT 2
T07123
nine-cis-epoxycarotenoid dioxygenase - tomato
N:Alternate names: probable neoxanthin cleavage enzyme
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07123
R:Burridge, A.
Submitted to the EMBL Data Library, January 1998
A:Reference number: Z15934
A:Accession: T07123
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-605 <BUR>
A:Cross-references: EMBL:Z97215; PIDN:CAB10168.1

Query Match 72.4%; Score 2280.5; DB 2; Length 605;
Best Local Similarity 71.7%; Pred. No. 6e-165;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;
QY 5 TATAAVSGRWLGNGHTOPPLSSOSSDLSVCS---SLPMASRVTRKLVSSALHTPALH 61
Db 3 TTTSHATNWI-----KTLKSPSKEFGFASNSISLLKNQHNROSLNINSLSQAPILH 57
QY 62 FPKQSSN--SPA--IVVKPKAESN-----TKOMNI.FQRAAAALDAEAGFLVSHKELHP 112
Db 58 FPKQSSNYQTPKNNTISHPKQENNNSSSTSKNVLQKAAAMALDAVESALTKELEHP 117
QY 113 LPKTADPSVOIAGNPAPVNEQVRNLPVVKLPDSIKGVYVNGANPLHEPVGTGHFFED 172
Db 118 LPKTADPRVQISGNFAPVNEQVRNLPVVKLPDSIKGVYVNGANPLHEPVGTGHFFED 177
QY 173 GDMVHAVKFEHGSASACRFTQTRNFVQEROLGRPVFPKALGELHGTGIALMLFYAR 232
Db 178 GDMVHAVQFKNGSASACRFTETERLYQEKALGRPVFPKALGELHGTGIALMLFYAR 237
QY 233 AAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVOITPNDGLTKVGRDFDGOLES 292
Db 238 GLFGLVDHSGKGVANAGLVYNGRLLAMSEDDLPYQVOITPNDGLTKVGRDFDGOLES 297
QY 293 TMIAHPKVDPESEGLFALSVDVSKPKYKPFSPDGKSPDVEIQLOOPTMHDFATE 352
Db 298 TMIAHPKLDPVSGELFALSVDVSKPKYKPFSPDGKSPDVEIQLOOPTMHDFATE 357
QY 353 NFVVVDDQVQVFKLPMIRGSGPVVYDKNKVARFGLDKYAEADSSNIKWDAPDFCFHL 412
Db 358 NFVVIPDQVQVFKMSEIRMGSGPVVYDKNKVSFRGILDKYAKGSDLKWVEVDFCFHL 417
QY 413 NNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 477
Db 418 NNAWEAEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGKSTRKAIENPDEQ 472
QY 473 QVNLEAGMVRNMLGRKTKFAYLALAEAPKPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPL 532
Db 478 QVNLEAGMVRNMLGRKTKFAYLALAEAPKPKVSGFAKVDLTGTEVKKHLYGDNRYGGPPL 537
QY 533 FLPGE--GGEDEGYILCFVHDEKTKWSELOITVNAVSELEATVKLPSPVPGPHGTGIGA 590
Db 538 FLPGE--GGEDEGYILCFVHDEKTKWSELOITVNAVSELEATVKLPSPVPGPHGTGIGA 597

Db 538 FLPRDPNSKEEDGYILAFVHDEKTKWSELOITVNAVSELEATVKLPSPVPGPHGTGIFIN 597
QY 591 ADDLAKQ 597
Db 598 ANDLANQ 604

RESULT 3
E96812
protein F3f9.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96812
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, S.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mailli, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96812
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <STO>
A:Cross-references: GB:AE005173; NID:g8052533; PIDN:AAF71797.1; GSPDB:GN00141
C:Genetics:
A:Gene: F3f9.10
A:Map position: 1

Query Match 68.9%; Score 2170.5; DB 2; Length 657;
Best Local Similarity 67.7%; Pred. No. 1.5e-156;
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;
QY 1 MASFTATAAVSQRWL-----GGNHTQPPPLSSOSSDLSVCSLSP-----MASRYT 45
Db 51 MASTILLPSTSTQFLDRTSTSSSRPKLOS-----LSFSTLRNKKLVVPCYVSSVYN 105
QY 46 RKLNVSSALHT---PPALHFFKQSSNSPAIVVKKAKESNTKQMLFORAAAAALDAE 101
Db 106 KSSVSSSSLSQSPTEPPSW---KKLCNDVTNLI-PKTTNQNK-LNPVQRTAMVLDVE 160
QY 102 GLVSHK-KLHLPKPTADPSVOIAGNPAPVNEQVRNLPVVKLPDSIKGVYVNGANP 160
Db 161 NAMI SHERRRHPKPTADPSVOIAGNPAPVNEQVRNLPVVKLPDSIKGVYVNGANP 220
QY 161 LHEPVTHGHHFFDGDGMVHAVKFEHGSASACRFTQTRNFVQEROLGRPVFPKALGELHGH 220
Db 221 LHKPVSGHHLPDGDGMVHAVKFEHGSASACRFTQTRNFVQEROLGRPVFPKALGELHGH 280
QY 221 TGIARLMIFYARAAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVOITPNDGLTK 280
Db 281 LGIAKLMFLNTRGLFGLVDPTGGLGVANAGLVYNGRLLAMSEDDLPYHVKVTQGDLET 340
QY 281 VGRFDGQLESTMTAHPKVDPESEGLFALSVDVSKPKYKPFSPDGKSPDVEIQLO 340
Db 341 SGRYDFDQGLKSTMTAHPKVDPESEGLFALSVDVSKPKYKPFSPDGKSPDVEIQLO 400
QY 341 OPTMMHDFAITENFVVVDDQVQVFKLPMIRGSGPVVYDKNKVARFGLDKYAEADSSNIK 400
Db 401 OPTMHDFAITENFVVIPDQVQVFKLPMIRGSGPVVYDKNKVARFGLDKYAEADSSNIK 460
QY 401 WIDAPDCFCFHLNNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGTE 460
Db 461 WIEVPDCCFHLNNAWEPEETDEVVVIGSCMTPPDSIFNECDGLKSVLSEIRLNLTGTE 520
QY 461 STRPPLISNEQVQVFKLPMIRGSGPVVYDKNKVARFGLDKYAEADSSNIK 520
Db 521 STRPPLISNEQVQVFKLPMIRGSGPVVYDKNKVARFGLDKYAEADSSNIK 577

QY	521	LYGDNRRYGGPFLPLPGEGGEDGYILCFVHDSKTKSELQIVNAVSLVEATVKLPSRV	580
Db	578	IYGEGKYGGPFLPGSDGEDGYIMFVHDESEKYLQINAVNMKLEATVTLPSPV	637
QY	581	PYGFHGTFIGADDLAKQVV	599
Db	638	PYGFHGTFISKDLSKOAL	656

RESULT 4
A86425
probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86425
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Chin, C.W.; Hughes, B.; Huizar, L.
Nature 408, 815-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86425
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-589 <STO>
A:Cross-references: GB:A005172; NID:g11094779; PIDN:AAG29711.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 54.9%; Score 2044.5; DB 2; Length 589;
Best Local Similarity 68.1%; Pred. No. 4.7e-147;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

QY	27	QSQSDLSYCSSSLPMASVRTRKLNVSALHTPPALHPKOSSSPAIVTKPKAKESNTKOM	86
Db	27	SPSSSYFNTKPR---RRKLSANSVSOTNLLNFFNPSNPPII-----PEKDTSRW	76
QY	87	NLFQRAAALDAAEGLFVSHKELHPLPKTADFSVOIAGFAPNVEQPVRNRLPVVGKLP	146
Db	77	NPLQRAASAALOFAETALLRRERSKPLPKTVPDRHQISGNYPAPVEQSVKSSLSVDGKIP	136
QY	147	DSITKGVVYRNGANPLHEPVTHGHEFDGGMHVAHFHSGSASACRFQTIRNFVQEROLG	206
Db	137	DCIDGYVLNGANPLFEPVSGHLLFDGGMHVAFKITNGDASYSCRFETETPLKVEKOLG	196
QY	207	RPVFPKAIAGELHGHGIAKIMLFYARAAAGIVDPAHGTVGANAGLVYFNGRLLAMSEDDL	266
Db	197	SPILPRAIGELHGHGIAKIMLFYARGLFGLLNHRNGTGVANAGLVYFHDRLLAMSEDDL	256
QY	267	PYQVQLTPNGDLKTVGRFDFDQOLSEMTIAHPKVDPESCELFALSYDVVSKPYLKYFPES	326
Db	257	PYQVRVTDNGDLDTGRFDFDQOLSSAMTAHPKIDVTKELFALSYDVVSKPYLKYEFKS	316
QY	327	PDGTSPPVEIQLDQPTMMHDFAITENFVVDPQOVVKFLPEMIRGGSPVYVDKNKYARF	386
Db	317	PEGEKSPVEIPLASPTMMHDFAITENFVDPQOVVKFLSDMFLGKSPVYIDGKISR	376
QY	387	GILDKYAEDSSNTKWIADPCFCFLHNAWABEPDETVVWIGSCMTPPDSTFNESDENLK	446
Db	377	GILPRAKADASEMWVWESPETECFHLNAWESPETEDVVVIGSCMTPADSIFNECDQOLN	436
QY	447	SVLSEIRLNLTGSTRRPTISNEDQQVNLEAGVYWRNMLGRKTKFAYLALAEPPKVS	506
Db	437	SVLSEIRLNLTGKSTRRTIIP-SVQMNLEAGVYWRNMLGRKTRYALALAEPPKVS	495
QY	507	FAKVDLITGEVAKKHYLDNRXGGEPFLTPG--EGGEDEGYILCFVHDEKTKWSELQIVN	564
Db	496	FAKVDLSTGEVAKKHYPYGGKYGGEFFPLPRGLEDGEDDGYINGFVHDESWESSELHVN	555

QY 565 AVSLEVEATVKLPSRVPGHGHTFGAGDDLAQ 597
 |||:::|||||::| | : |
Dd 556 AVTLEAEATVKLPSRVPGHGHTFVNSADMLNQ 588

RESULT 5
T04531

nine-cls-epoxy-carotenoid dioxygenase homolog F28J12.10 - *Arabidopsis thaliana*
 N;Alternate names: hypothetical protein T9A21.200
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C;Accession: T04531; T04937

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A. submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377

A;Accession: T04531
A;Molecule type: DNA

A;Residues: 1-583 <BEV>

A;Cross-references: EMBL:AL021710

A: Experimental source: Cultiyar Columbia: PAC clone F39113

R: Bevan, M.; Murphy, C.; Ridgway, R.; Hughes, C. *Plant* 1994, 15: 541-544.

R; Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, R.

submitted to the Protein Sequence Database, February 1999

A; Reference number: Z15390

A;Accession: T04937

A;Molecule type: DNA

A;Residues: 1-377 <BEW>

A;Cross-references: EMBL:AL021713

A: Experimental source: cultivar Columbia: PAC clone 40331

A; experimental source: cultivar Columbia; BAC clone T9A21
C. Genetics.

C; Genetics:
A: Man position. A

A;Map position: 4

A;Note: F28J12.10; T9A21.200

Query Match 63.2%; Score 1991; DB 2; Length 583;
Best Local Similarity 64.5%; Pred. No. 5.3e-143;
Matches 380; Conservative 82; Mismatches 105; Indels 2

QY 16 GCNHTQPTLSSQSSQSLV---CSSLPMSRVTRKLVSSALHTPPALHFPKQSSNSPAIV 73

Db 11 GGKLTWP----QAQIDLGRPIKQPKVIKCTVQIDVTE-LTKKRQLFRTTATPP--- 62

```

Qy 74 VPKAKESNTKOMNLFQRAAAAALDAEGLVSHKHLPLPKTADPPSVQIAGNFAPVNEQ 133
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 -----QUNPLRNTFOKAAAIAIDAERALISHEOSSLPKTADPPRVQIAGNVSVPFES 116

```

QY 134 PVRNLPVVGKLPDSIKGYVVRNGANPLHEPVTGHHFFDGDGMVHAVKFEHGSASYACRF 193

Db	117	SVRRNLTVEGTIPDCIDGVYIRNGANGPMFEP	TAGHHLFDGDMVHVKITNGSASYACRF	176
QY	194	TQTNRFFVQERQLGRPVFPKAIGELHGHGTGI	ARLMIFYARAAAGIVDPDPAHGTGVANAGLVY	253

```
Db      177  TKTERLVQEKRLGRPVFKAIGELHGHSGIARLMFYARGLCGLINNQNGVGVANAGLVY 236
```

```

Qy      254 FNGRLAMSEDL PYQVQITP NGDLKTVGRFDG QGLSTMI AHPKVD PESGELFALSVD 313
          || ||||| ||||| :: || ||||| ||||| : ||||| : || |||||
Db      237 FNNRLAMSEDL PYQLKITQT GDLQTVGRYDFD GQLKSAMIAH PKLDPTVKELHALSVD 296

```

QY 314 VVSKPYLKYFRFSPDGTGKSPDVEIQLDQPTMMHDFATENFVVVPDQVVFKLPEMIRGG 373

Db	297	VVKPYLKYFRSPDGKVSPELEIPLPTMIHDAITENFVVPDQQVVKLGEMISK	356
QY	374	SPVVYDKNKVARFGIILDKYAEEDSNIKWIDAPDCFCFHLWNAREPFPETDEVVVIGSQMTP	433

db 357 SPVFDGKVSRLGINKPKDATEASQIIWNSPETFCFHLWNAMESPETEEIIVIGSCNSP 416

QY 434 PDSIFNESDENKLSVLSERLNLTGSTRPPIISNEDQQVNLEAGMVNPNNLGRKTFFA 493
||||| ||||| :||||| :||| ||||| ||||| ||||| |||||
db 417 ADSIENRDESLRSVSEIRNLTRKTPPSII VNRD - VAN ETGMANDY GRSDND 474

QY 494 YLALAEWPVKVSGFAKVDLTITGEVKKHLYGDNRYGCEPLFLPCEGG---EEDEGYILCF 549

Db 475 FLAIYPPKVSQFAKVDLCCTGEMKKYIYGGEKYGGEFFFLPNSNGNGEENEDDGYIFCH 534

QY 550 VHDEKTKWSELOIYNVAVSLEVEATVKLPSRVPYGFHGTGICADDLAKOV 598
 Db 535 VHDETKTSELQIINAVNKLKLEATIKLPSRVPYGFHGTIVDSNELVDOL 583

RESULT 6
 T04351
 N:Alternate names: vpl4 protein - maize
 C:Species: Zea mays (maize)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
 C:Accession: T04351
 R:Tan, B.C.; McCarthy, D.R.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z15304
 A:Accession: T04351
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-604 <TAN>
 A:Cross-references: EMBL:U95953; NID:g2232016; PIDN:AAB62181.1; PID:g2232017
 A:Experimental source: strain W22
 C:Genetics:
 A:Gene: vpl4
 A:Map position: 1
 C:Function:
 A:Description: catalyzes oxidative cleavage of 9-cis-epoxy carotenoids to yield xanthoxi

Query Match 61.3%; Score 1930; DB 2: Length 604;
 Best Local Similarity 63.3%; Pred. No. 2.4e-138;
 Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PPLSSSSDLSYCSLPMASRVTRKLNVSALHTP---PALHPFKOSSNSPAIVV--- 74
 Db 17 PARSRARAS-----NSVRFSRAVSSVPPABCLQAPFKPVADLPAPSRKPAALVAPGHA 71

QY 75 -KPKAKESNTKOMNLFORAAAALDA-AEGLVLS-HEKLHPLKPTADPSVOIAGNAPAVN 131
 Db 72 AAPRKAEGGKQLNLFORAAAALDAEEGFVANVLERPHGLPSTADPAVO-AAGNEAPVG 131

QY 132 EQPVRRRLVGVGKLPDSTIKGVYVRNGANPLHEPVTHGFFDGGMVHAKVEHSA-SYA 190
 Db 132 ERPPVHELVPSSRTPPFIDGVYARNGANPCDPVAGHLLFDGDMVHALRNGAAESYA 191

QY 191 CRFTQTRNFVOEROLGRVPFKAIGELHGTGIALMLFYARAAAGIVDPAHGTGVANAG 250
 Db 192 CRFTETARLKOERAIGRPVFPKALGELHGHSGIARLALFYARAAAGLVDPASAGTVANAG 251

QY 251 LVYFNGRLAMSEDDLPOVQOITPNGDLKTVGRFDGQLESTMIAPKVPDPESGELFAL 310
 Db 252 LVYFNGRLAMSEDDLPHVRAVADGDDLETGRYDFDGLGCAMIAHPKLPDPAATGELHAL 311

QY 311 SYDVVSKPYLYKFRFSPDGTSPDVEIQDQPTMMHDFAITENFVVPDQOVWFKLPEMI 370
 Db 312 SYDIKRPYLYKFRFSPDGTSPDVEIQDQPTMMHDFAITENFVVPDQOVWFKLPEMI 371

QY 371 RGSFPVYDKNKVARFGLLDKYADESSNIKIDAPDPCFHLNNAWEEPETDEVVVTGSC 430
 Db 372 RGSFPVYDKNKVARFGLLDKYADESSNIKIDAPDPCFHLNNAWEEPETDEVVVTGSC 431

QY 431 MTPDSTIFNESDENLKSLSIRNLKGTSTRRPITSNEDQOVNLEAGYKRNMLGRKT 490
 Db 432 MTPADSTIFNESDENLKSLSIRNLKGTSTRRPITSNEDQOVNLEAGYKRNMLGRKT 490

QY 491 KFAYLALAEAPWPKVSGFAKVLDITTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 545
 Db 491 RYAYLALAEAPWPKVSGFAKVLDITTEVKKHLYGDNRYGGEPLFLPGEGE-----EDEGY 550

QY 546 ILCFVHDEKTKWSELOIYNVAVSLEVEATVKLPSRVPYGFHGTGICADDLAKOV 597
 Db 551 VLTFVHDERAGTSLLVVVNAADIRLEATVQLPSRVPYGFHGTGICADDLAKOV 602

RESULT 7

T04438
 hypothetical protein T18B16.140 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04438
 R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansgore, W.; Bancroft, I.; Mewes,
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15359
 A:Accession: T04438
 A:Molecule type: DNA
 A:Residues: 1-595 <BEV>
 A:Cross-references: EMBL:AL021687
 A:Experimental source: cultivar Columbia; BAC clone T18B16
 C:Genetics:
 A:Map position: 4
 A:Note: T18B16.140

Query Match 30.7%; Score 968; DB 2: Length 595;
 Best Local Similarity 36.7%; Pred. No. 2.5e-65;
 Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLNVSALHTPPTALHPFKOSSNSPAIVVKKAKESNTKOMNLFORAAAALDAAE 101
 Db 42 SPITNPSSDNDNRNRPKPTLH---NRTNHTLVSSPPKLRPEMTLALF---TTVEDVIN 94

QY 102 GFLVSHEKLHPLKPTADPSVOIAGNEAPVNBQPVRRNLPPV-GKLPDSIKGVYVRNGANP 160
 Db 95 TFIDP-----PSRPSVDPKHVLSDNFAPVLDLPPTDCEIHTGLTSLNGAYIRNGPNP 149

QY 161 LHEPVTHGFFDGGMVHAKVEHGSASACRFTQTRNFVOEROLGRVPFKAIGELHGH 220
 Db 150 QFLPRGPHVLLFDGDMHAIKHNGKATLCRSYKTVKYVKEKOTGAPVMPNVFSGENG 209

QY 221 T-GIARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLPOVQOITPNGDLK 279
 Db 210 TASVARGALTAAARVITGQYNPVNGIGLANTSLAFPSNRFLFALGESDLFYAVRLTESGDIE 269

QY 280 TVGRFDDGQLESTMIAPKVPDPESGELFALSVDWVSKPYLYKFRFSPDGTSPDVEI-Q 338
 Db 270 TIGYDFDGLKAMSTAHPTDPTGTETFAFRYGPV-PPFLTYFRFDSAGKKQDVPLFS 328

QY 339 LDQPTMMHDFAITENFVVPDQOVVFX---LPEMIRGSGPVVYDKNKVARFGLDKYAE 395
 Db 329 MTSFSLHDFAITKRAHFAIBIQLMRNMMLDLVLEGGSPVGTDNKGTTPRLGVIPKYAGD 388

QY 396 SSNTKWDADPCFCFHLNNAWEEPETDEVVVTGSCMPPDPIFNESDENLKSLSIRLN 455
 Db 389 EEMKWEVPGFNIIHAINAWDDGNSVLIAPNISIEHTLERMD-LVHALVEKVKID 447

QY 456 LKTGSTRRRPITSNEDQOVNLEAGYKRNMLGRKTFAYLALAEAPWPKVSGFAKVLDITTE 515
 Db 448 LVTGIVRRHPISAR-----NLDEAVINPAFLGRCSRYVYAAIGDPMKPGISGVKLDVSKG 502

QY 516 E-----VKKHLYGDNRYGGEPLFLPGEGE-----BEDGYYILCFVHDEKTKWSELOIYNVAV 566
 Db 503 DRDQCTVARRMYSGCYGGEFFVARDPDPNPEAEDDGGVYVTVHDEVTGESKFLVMDAK 562

QY 567 S--LEVEATVKLPSRVPYGFHGTGICADDLAK 596
 Db 563 SPELEIVAAVRLPRVPYGFHGLFVKESDLNK 594

RESULT 8
 T49193
 neoxanthin cleavage enzyme ncl - Arabidopsis thaliana
 N:Alternate names: protein MAA21.150
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C:Accession: T49193
 R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25018
 A:Accession: T49193

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-538 <RIE>
A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150
A:Experimental source: cultivar Columbia; BAC clone MAA21
C:Genetics:
A:Gene: ATSP:MAA21.150
A:Map position: 3
A:Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 49; Gaps 13;

QY 64 KSSNSPAIVVVKPK-AKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIIVHPRSKGFSKLLDLLELVVKLM-----HDASLPLH-----Y 47
QY 123 IAGNFAPV-NQGPVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHHFFDGDGMVHAVK 181
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYVHFFDGDGMHIGVR 107
QY 182 FEHGSASACRFTQTNRFVQERQGRPVPPKAIGELHGHGTGIALMLFYARAAAGIVDPA 241
DB 108 IKDKGATVSVRYVKTSLKQEEFFGAAKFMK-IGDLKGFGLLMVNVQQLRTKLKILNT 166
QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRDFDGDGOLSTMIAPKVD 301
DB 167 YNGNTANTALVYHKGKLLALQADKPYVVKVLEDGLQTLIGIDYDKRLTHSFTAHPKVD 226
QY 302 PESGELFALSVDVSKPYLKYFRSPDGTSPDVETIQDPTMMHDFAITENFVVPDQ 361
DB 227 PVTGEMTFEGYS-HTPPYLYTVRISKDGMHDPVITISEPIMMHDFAITETIYAFMDLP 285
QY 362 VVFKLPEMIRGSPV-YDKNKVARFGLDKYAESSNIKWIDAPDCFCFHLWNAWEEPE 420
DB 286 MHFRPKEMVKKMIYSPDPTKKARFGLVPRYAKDELIRWELPNCFIHNNANWE--E 343
QY 421 TDEVVVIGSCMTPD-----SIFNESDENLKSVLSEILNLTGSTRPPIISNEDQOVN 475
DB 344 EDEVVLTICRLNFDLDMVSGKVKLENGELNEMFNKMGTSASOKLSASA----- 398
QY 476 LEAGVNRNMLGRKTKFAYLALAPWPKVSGFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTYTKKORYVYGTILDSIAKVTGIIFDLHBAEATGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPGEGEDEGYILCFVHDEKTKWSELQIVNAVYSLEVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPRETAEDDGYLFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGFHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLOEQTL 537

RESULT 9
T51734
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51734
R:Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.
J. Exp. Bot. 49, 1893-1894, 1998
A:Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture
A:Reference number: Z24454
C:Accession: T51734
A>Status: preliminary; translated from GH/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-538 <NEI>
A:Cross-references: EMBL:AJ005813; PIDN:CAA06712.1
A:Experimental source: Cultivar Landsberg erecta
C:Genetics:
A:Gene: ncl

Query Match 29.7%; Score 937; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 4.8e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KSSNSPAIVVVKPK-AKESNTKOMNLFQRAAAALDAAGFLVSHKHLPLPKTADPSVQ 122
DB 4 KLSGSSIIIVHPRSKGFSKLLDLLELVVKLM-----HDASLPLH-----Y 47
QY 123 IAGNFAPV-NEOPVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHHFFDGDGMVHAVK 181
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYVHFFDGDGMHIGVR 107
QY 182 FEHGSASACRFTQTNRFVQERQGRPVPPKAIGELHGHGTGIALMLFYARAAAGIVDPA 241
DB 108 IKDKGATVSVRYVKTSLKQEEFFGAAKFMK-IGDLKGFGLLMVNIQQLRTKLKILNT 166
QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRDFDGDGOLSTMIAPKVD 301
DB 167 YNGNTANTALVYHKGKLLALQADKPYVVKVLEDGLQTLIGIDYDKRLTHSFTAHPKVD 226
QY 302 PESGELFALSVDVSKPYLKYFRSPDGTSPDVETIQDPTMMHDFAITENFVVPDQ 361
DB 227 PVTGEMTFEGYS-HTPPYLYTVRISKDGMHDPVITISEPIMMHDFAITETIYAFMDLP 285
QY 362 VVFKLPEMIRGSPV-YDKNKVARFGLDKYAESSNIKWIDAPDCFCFHLWNAWEEPE 420
DB 286 MHFRPKEMVKKMIYSPDPTKKARFGLVPRYAKDELIRWELPNCFIHNNANWE--E 343
QY 421 TDEVVVIGSCMTPD-----SIFNESDENLKSVLSEILNLTGSTRPPIISNEDQOVN 475
DB 344 EDEVVLTICRLNFDLDMVSGKVKLENGELNEMFNKMGTSASOKLSASA----- 398
QY 476 LEAGVNRNMLGRKTKFAYLALAPWPKVSGFAKVD-----LTTGEVKKHLY-- 522
DB 399 VDFPRINECTYTKKORYVYGTILDSIAKVTGIIFDLHBAEATGKRMLEVGNGIKGIYDL 458
QY 523 GDNRYGGEPLFLPGEGEDEGYILCFVHDEKTKWSELQIVNAVYSLEVE--ATVKLPSRV 580
DB 459 GEGRYGSEAIYVPRETAEDDGYLFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518
QY 581 PYGFHGTFIGADDLAKQVV 599
DB 519 PYGFHALFVTEEQLOEQTL 537

RESULT 10
AG1944
hypothetical protein all1106 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AG1944
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri:
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabat
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <KUR>
A:Cross-references: GB:HA000019; PIDN:BA073063.1; PID:gl7130452; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1106

Query Match 23.7%; Score 745.5; DB 2; Length 475;
Best Local Similarity 34.5%; Pred. No. 1.3e-48;
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

QY 123 IAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLHEPVTHGHHFFDGDGMVHAVK 182
DB 48 LSGNFAPIRDETTPPVKDLPHVGHFLPECLNGEFVRGPNPKFDVAGYVHFFDGDGMHIGVR 107

Db 22 LEGNAPVHEELTTDLAKVIGELPELSSMEVRNCPNFWPTIGQYHWFED3DGMHGVRI 81
QY 183 EHGSSASYACRTQTNNRVOEROLGRPVPPKALGELHGTGTLARMLFYARAAAGIVD--- 239
Db 82 SNGKATYNNRYVTRQWQIEHAGQAIW-----IGLMEPPQ 117
QY 240 ---PAHGTGVANAGVYFNGRLAMSEDDLPVOVITPENGDLKTVGDFDQLESTMA 296
Db 118 TELPSRNTG--NTALIWHAGQALLWESGAPYAIQVP---DLASIGETYNNQLSSAF 172
QY 297 HPKVDPESEGFALSVDVYVSKYLYKFREFSPDGTSPDVEIQDQOPTMMHDAFATENFV 356
Db 173 HPKVDPEVTGEMFFOYS--FAPPYLHYVVSATGELVTVPIIDLPMGVMMHDAFATANY 231
QY 357 VPDQVQVFKLPMIRMGSPVYDKNKVARFGILDYXEDSSNIKWIDAPDCCFHLWNAW 416
Db 232 FMDLPLTFVSERMQRGPEMLFESDRPSRFGILPRHG--DNSQIRWFEPAPSCVFTHLNAY 290
QY 417 EPEPTEVVWIG-----SCMTPPDSIFNESDENLK--SVLSEIRLNKLTGSTRPILSN 469
Db 291 E--DKDEVVLFACPMRSTIVLASPPS---QTDPEADIPRLHWRHFKLTGKVHEEML--- 342
QY 470 EDQVNNLEAGVYNNRMLGKRTKAYLA--LAE--PMKVSFGFAKVDLTGTGKKHLYGDNRY 527
Db 343 --DDVASEPRINENELGTOGTGYTSRLAKGSIFLFGILKYDLSNAKSNQYEGCGRY 400
QY 528 GCEPLFLPGEGE--EDEGYILCFVHDEKTKWSELOLVNNAVSLEVE--ATVKLPSRVPGF 584
Db 401 GSEAVFVRPGATVEDDGLWLIYVYDTEGESSELVYVNAQDINSEPIARVILPQRPVPGF 460
QY 585 HGTFIGADDL 594
Db 461 HGIWVTEQL 470

RESULT 11
T17019
Query Match 16.8%; Score 528.5; DB 2; Length 483;
Best Local Similarity 30.9%; Pred. No. 4.1e-32;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;
C:Species: Malus domestica (apple tree)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T17019
R:Watillon, B.; Kettmann, R.; Arredouani, A.; Hequet, J.; Boxus, P.; Burny, A.
Plant Mol. Biol. 36, 909-915, 1998
A:Title: Apple messenger RNAs related to bacterial lignostilbene dioxygenase and plant
A:Reference number: Z18655; MUID:98179104; PMID:9520281
A:Accession: T17019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-446 <WAT>
A:Cross-references: EMBL:Z93765; NID:g2924324; PIDN:CAB07784.1; PID:g2924325
A:Experimental source: cv. McIntosh, strain Wj/cik

Query Match 23.7%; Score 745; DB 2; Length 446;
Best Local Similarity 39.8%; Pred. No. 1.3e-48;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;
QY 231 ARAAGIVDPAGHTGVANAGLVYFNGRLAMSEDDLPYQVITPNGDLKTVGDFEDGQL 290
Db 76 ARVLTGOYNPANGIGLANTSLAFFGDRLYALGESDLPYSLRLTNSGDIETLGRHDFDGL 135
QY 291 ESTMIHAKVPDPSEGFALSVDVSKYLYKFREFSPDGTSPDVEI--QLDQPMHMDFA 349
Db 136 SMNTAHPKIDPDTGEAFAPRYGFI--RPFLTYREFDSNGYKQDPVIFSMVTTFELHDA 194
QY 350 ITENFVVPDQVQVFKLPEMI--RGGSPVYDKNKVARFGILDKYAESSSNIKWIDAPDCF 408
Db 195 ITRKHAIFADIQGLNIDMLTKRATPFGDLPKVPRIQVPIYAKDESMRWFEPGFN 254
QY 409 CFHLWNAWEPEPTEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRPILSN 468
Db 255 GVHATNWD--EDDAIVWAPNVLSAEHVLERVD--LVHCLVEKVRIDKLTGIVTRQPIST 311
QY 469 NEDQVNNLEAGVYNNRMLGKRTKAYLALAEPMKVSFGFAKVDLTGGE-----VKKHLYG 523

Db 312 R-----NLDFAVINFAYLGRKNKYVYAAEGDPMKISGVKWLDSNVHEKCEIVASRMFG 366
QY 524 DNRVGCEPLFLPGE-----GGEDEGYILCFVHDEKTKWSELOLVNNAVS--LEVEATVKLP 577
Db 367 PCGYGCEFFVAREPENPEADENGLFVSYVHDEKAGESRFLVMDAKSPQDLIVAARMP 426
QY 578 SRVPYGFHGTIFIGADDL 594
Db 427 RRVPGFHLFVRESDL 443

RESULT 12
E87345
Query Match 16.8%; Score 528.5; DB 2; Length 483;
Best Local Similarity 30.9%; Pred. No. 4.1e-32;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87345
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, R.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U S A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AE005673; NID:gl3422015; PIDN:AAK22761.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0776

QY 108 EKLHPLPKTADPS--VOIAGNAPVNEQFVRNLPV--GKLPDSIKGVVVRNGANPLHEP 164
Db 4 ERLPPVRTSLGPTNHPYMTGPTPQHEEVNAMDLEVGAIADLGDGVYLRNTEPVHDP 63
QY 165 VTGHHFDFDGDGMVHAVKFEHGSASYACRTQTNNRVOEROLGRPVPPKALGELHGTGIA 224
Db 64 IGRYHFFDGDGMHIOEFKGGATYNNRVRTRCFEAEQFVNEGLW---GGLMDGPGVS 119
QY 225 RMLFYARAAAGIVDPAHGT--GVANAGLVYFNGRLAMSEDDLPYQV---VOITPNGDL 278
Db 120 K-----RPGFG---AHGALKDSASTDIVHNGEAIATF-----YQCGEAYRLDPL--TL 163
QY 279 KTVGREFDEQLESTMIHAKVPDPSEGFALSVDVSK--PYLYKFREFSPDGTSPDVE 336
Db 164 ENLGVASW--APLEG--VSAHPKVDTEATGELMFFNY---SKAWPYMHYGVVPGDKRKYQG 218
QY 337 IQLDQPTMHDFAITENFVVPDQVQVFKLPEMIRGSPVYDKNKVARFGILDKYAES 396
Db 219 VPLPGRLPHDMAFSSKYAILNDLPVFWQELMARDIHAVRLHKGIPSRFALVPR---EG 275
QY 397 SNTKWIDAPDCFHHLWNAWEPEPTEVVIGSCMTPP-----DSIFNESDEN 444
Db 276 GEPRWEAEPTYYLHMLNAYE--DGDEVLDGYEQEKPIRPLEGAPDGHGLMAYLDEH 333
QY 445 LKSVLSEI---RLNIAKTGSTRPILSNEDQVNNLEAGVYNNRMLGKRTKAYLALAE 500
Db 334 --SFLPKLHWRNLTGTEKHL-----DDRV--LFGFMFNQKYAKPRYAIYSTAKPG 386
QY 501 WPKVSGFAKVDLTGTGEVKKHLYGDNRYGGEPLFLPGE--EEDGYILCFVHDEKTKWSE 559
Db 387 WFLNFGVFKHDLTGESWSIALPEGYASAPAPKVGAVDEDDGYLVSLIDENKASE 446
QY 560 LQIVNNAVSLEVEATVKLPSRVPGFHTGTFIGADDLAK 596
Db 447 CLIVDAKRFVWCRIALPHKLSSTGTHVWAGREMLTK 483

Nature 393, 537-544, 1998

A;Residues: 1-490 <RAN>
A;Cross-references: EMBL:D0914; GB:AB001339; NID:g1653477; PIDN:BAAL8428.1
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, Jun

Search completed: May 19, 2003, 15:49:09
Job time : 25 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:14 ; Search time 27 Seconds
(without alignments)
920.160 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAAVSGRWLGNGHT.....VPYGRHFGIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120	3.8	616	1 NXFL_COTJJA	P58797 coturnix co
2	114	3.6	870	1 YCS0_YEAST	P25623 saccharomyc
3	113	3.6	466	1 DNAA_PROMI	P22837 proteus mir
4	107.5	3.4	658	1 PAK1_SCHPO	P50527 schizosacch
5	106	3.4	3562	1 PGCV_CHICK	Q90953 gallus gall
6	105	3.3	783	1 CAD5_MOUSE	P55284 mus musculu
7	105	3.3	2504	1 FAS_HUMAN	P49327 homo sapien
8	104.5	3.3	1079	1 APCE_FREDI	P16566 fremyella d
9	103	3.3	775	1 SYFB_AQUAE	C67620 aquifex aeo
10	102	3.2	2493	1 CYAA_USTRA	P49606 ustilago ma
11	101.5	3.2	520	1 GLGS_BRANA	Q9m462 brassica na
12	101.5	3.2	565	1 SCRL_SCHPO	Q14335 schizosacch
13	101	3.2	512	1 GLGT_VICFA	P52417 vicia faba
14	100	3.2	3770	1 ACVS_EMENI	P27742 emericella
15	99.5	3.2	383	1 PHVC_BACSU	Q31097 bacillus su
16	99.5	3.2	1256	1 ATIL_STAUI	P52081 staphylococ
17	98	3.1	741	1 MASZ_MYCTU	Q50596 mycobacteri
18	98	3.1	1788	1 YP72_CAEEL	Q05221 caenorhabdi
19	97.5	3.1	518	1 CP74_ARATH	Q96242 arabidopsis
20	97.5	3.1	691	1 SLP1_YEAST	P20795 saccharomyc
21	97	3.1	385	1 OXDC_BACSU	Q34714 bacillus su
22	97	3.1	636	1 KDPG_SCHPO	Q14019 schizosacch
23	96.5	3.1	426	1 UCR2_SCHPO	P78761 schizosacch
24	96.5	3.1	503	1 DUTA_BACSU	P39581 bacillus su
25	96.5	3.1	862	1 SMP2_YEAST	P32567 saccharomyc
26	96.5	3.1	3587	1 TYCB_BACBR	Q30408 b tyrocidin
27	96	3.0	462	1 DNAA_YERPE	Q82-9u7 yersinia pe
28	95.5	3.0	865	1 NRFA_PENUR	Q92269 penicillium
29	95.5	3.0	918	1 DNLL_RAT	Q91hy8 rattus norv
30	95.5	3.0	1144	1 DP3A_NEIMB	Q91x22 neisseria m
31	94.5	3.0	399	1 EX7L_CLOAB	Q97hd0 clostridium
32	94.5	3.0	590	1 HMAA_DROME	P29555 drosophila
33	94.5	3.0	638	1 GHR_RAT	P16310 rattus norv

34 94.5 3.0 778 1 RHP9_SCHPO
35 94.5 3.0 953 1 LKAB_PASHA
36 94.5 3.0 1046 1 RPOC_WEIHE
37 94 3.0 532 1 P72_MYCW
38 94 3.0 919 1 Y893_HUMAN
39 93.5 3.0 872 1 SYA_LACIA
40 93.5 3.0 1275 1 GNRP_HUMAN
41 93 3.0 3119 1 CALC_MOUSE
42 92.5 2.9 354 1 OTX1_HUMAN
43 92.5 2.9 746 1 PCAP_HUMAN
44 92.5 2.9 955 1 LKAA_PASHA
45 92.5 2.9 1057 1 SPSL_CITUN

ALIGNMENTS

RESULT 1

ID NXFL_COTJJA STANDARD; PRT; 616 AA.
 AC P58797;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).
 DE NXFL.
 GN Coturnix coturnix japonica (Japanese quail).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Coturnix.
 OC NCBI_TaxID=93934;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLU-246.
 RX MEDLINE=21257892; PubMed=11358864;
 RA Coburn G.A., Wiegand H.L., Kang Y., Ho D.N., Georgiadis M.M., Cullen B.R.;
 RT "Using viral species specificity to define a critical protein/RNA interaction surface.";
 RL Genes Dev. 15:1194-1205(2001).
 CC -!- FUNCTION: Involved in the nuclear export of mRNA from the nucleus to the cytoplasm.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are essential for the export of mRNA from the nucleus (By similarity).
 CC -!- MISCELLANEOUS: Changing a single residue, Glu-246 to Arg, the residue found in human NXFL, fully rescues both constitutive transport element (CTE) function and CTE binding.
 CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF343749; AAK58910.1;
 CC PROSITE: PS50177; NTF2_DOMAIN; 1.
 CC Transprot; mRNA transport; Nuclear protein; Repeat;
 CC Leucine-rich repeat; Multigene family.
 CC DOMAIN 116 195 RNA-BINDING (RRM) (BY SIMILARITY).
 FT REPEAT 263 288 LRR 1.
 FT REPEAT 289 312 LRR 2.
 FT REPEAT 313 340 LRR 3.
 FT REPEAT 341 368 LRR 4.
 FT DOMAIN 383 533 NTF2.
 FT DOMAIN 563 605 UBA-LIKE (BY SIMILARITY).

FT	MUTAGEN	246	246	Q>:RESCUES CIE FUNCTION AND BINDING.
SQ	SEQUENCE	616 AA;	69329 MW;	D14903DDCF1BFLBD CRC64;
	Query Match	3.8%;	Score 120;	DB 1; Length 616;
	Best Local Similarity	22.0%;	Pred. No. 0.29;	
	Matches	86;	Conservative	57; Mismatches 132; Indels 116; Gaps 18
QY	282	GRFDFDQLESTM-----	IAHPKVDPSGELFALSVDVWSPKLYKIFRSPD	328
DB	28	GRGPFGRKMYSEMNRNSRRNGTGTSSSRQRODEDGDV-ANS-DAHADAPRGRIYLPYGR	85	
QY	329	GTKSPDVEITLDQPT-----	MMHDFAIT	360
DB	86	PSRAHLIVRRDLPPDRSGGSRDGRNRNWFKITIPYKGYDKTLLWLISSIQNLCSVPPT	145	
QY	361	QVVFKLPEMIRGSPVVDYKKNVARGILDKVAEDSSNIKWIDAPDCFCFLHNNWEPE	420	
DB	146	PVEFH-----	YDHNK-AQFYVD--ATTASALKQVS-----	180
QY	421	TDEVVVIQSWTTPDSITN-----	ESDENLKSVL-----	469
DB	181	NKVVIIIIINSSAPPFSLONELKPEIEQLKVCMSKRYDGAQALDLK-CLRVDPDLVSQS	239	
QY	470	-----	EDQOVNLEAGMVNENMLGRKTKFAYLALAEWPVKVSGFAKD	511
DB	240	IDVVLNQSRCHMMVLRITIENTPELQSLNLSNKKLYRLDDLAELAI-----	KAAGLKVLD	294
QY	512	LTTGEVKKHLYGDNRYG--GEPLFLPGE-----	GGEDEGYILCFVHDEKTKWSEL---	563
DB	295	LSRNLKSERELDKVGLKLEELWDGNPLCDGFRDQSSYSSVRRFPKLLRIDGHLP	354	
QY	564	NAVSLEVEATVKLPSPVYGFHGTIGADDL	594	
DB	355	PRIAEDVAPVTLPP-----	CKGSYFGSDDL	380
RESULT 2	YCS0_YEAST	STANDARD;	PRT;	870 AA.
ID	YCS0_YEAST	STANDARD;	PRT;	870 AA.
AC	P25623;	P25622;		
DT	01-MAR-1992	(Rel. 22, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
GN	Hypothetical 96.1 kDa protein in RIM1-RPS14A intergenic region.			
DE	YCR030C OR YCR30C/YCR29C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NB1_TaxID=4932;			
ON	[1]			
RF	SEQUENCE FROM N.A.			
RA	Cederberg H., Hohmann S., Schaeff-Gerstenschlaeger I., Huse K.,			
RA	Zimmermann F.K.;			
RL	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	REVISIONS.			
RA	Gromadka R.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: SOME, TO S.POMBE SPEC4C3.06.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL;	X59720; CAA42322.1;			
DR	DR			
DR	SGD; S0000626; YCR030C.			
KW	Hypothetical protein.			
SQ	SEQUENCE 870 AA; 96125 MW; 3F2CE2F15625277 CRC64;			

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58352; AAA83958.1; -
CC PIR: J00733; IOEBV
CC InterPro: IPR001957; Bac_DnaA.
CC Pfam: PF00308; bac_dnaA; 1.
CC PRINTS: PR00051; DNaA.
CC TIGRfam: TIGR00362; DnaA; 1.
CC PROSITE: PS01008; DNaA; 1.
CC DNA replication; DNA-binding; ATP-binding.
CC NP_BIND 171 178
CC FT NP_BIND 171 178
CC SQ SEQUENCE 466 AA; 52974 MW; F8B67C142FE9FA41 CRC64;
CC -----
CC Query Match
CC Best Local Similarity 31.6%; Score 113; DB 1; Length 466;
CC Matches 84; Conservative 48; Mismatches 112; Indels 140; Gaps 21;
CC
CC QY 105 VSHEKLHPLKT-ADPSVOIAGNAPVNEQPVRRNLVVYVGGKLPDS-IKGVYVRNGANPLH 162
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 83 VSARTIESVPKTVTHPAV-----NSTFTNSQPVYR---PSWDNQPSQLPELNYSNVNPKH 135
CC QY 163 EPVTGHHFFDGDGMVHAKFEGHGSYACRFOTNRFCVEROLG--RPVFPKAIAGELGH 220
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 136 K-----FD-----NFVEGKSNQLAR--AAARQVADNPGGAYNPLF-----LYGG 172
CC QY 221 TGIAKMLFYARAAAGVDPAGTGV---ANAGLVYFNGRLLAMSEDDLPYQVQIITPN 276
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 173 TGLGTHLLH-----AVGNSIMERKANAKVVMYH-----SERVQDMVKALQNN 216
CC QY 277 DLKTVGRDFDQGLRSTMIAPKVPDSEGEFALSVDVVVSKPLKYFRPSDGTSPDVE 336
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 217 ALE-----DFK-----RYR-SVDAL-IDDIQ 237
CC QY 337 TQLDPTMMHDPATFNFWVPDQWVF---KLPMIRGGSPVVDKKNVAREFGILDXYA 393
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 238 PFANKRSQEEFFHFNALLGNQIILTSRYPEING-----VEDRLK-SRFG-----286
CC QY 334 EDSSNIKIDAPDCFCFHLNWAEEPEPDEVVVGSCMTTPDPSIFNESDENLKSVLSEI- 452
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 287 -----WGLTVAIEPPELETRVAI-----LMKKADENQIQLPDEVA 321
CC QY 453 -----RLNLKTGSTRRPIISN 469
CC Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 322 FFIARLRSNVRELEGALNRVIAN 345
CC -----
CC RESULT 4
CC PAK1_SCHPO
CC ID PAK1_SCHPO STANDARD; PRT: 658 AA.
CC AC P50527;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 30-JUN-2000 (Rel. 39, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Serine/threonine-protein kinase pak1/shk1 (EC 2.7.1.-).
CC GN PAK1 OR SHK1 OR ORB2 OR SPBC1604.14C.
CC OS Schizosaccharomyces pombe (Fission yeast).
CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC OC Schizosaccharomycetes.
CC OX NCBU_TaxID=4896;
CC RN [1]
CC SEQUENCE FROM N.A.
CC RX MEDLINE=96112805; PubMed=8846783;
CC RA Ottilie S., Miller P.J., Johnson D.I., Creasy C.L., Sellis M.A.,
CC RA Bagrodia S., Forsburg S.L., Chernoff J.,
CC RT "Fission yeast pak1+ encodes a protein kinase that interacts with

```

```

RT Cdc42p and is involved in the control of cell polarity and mating.";
RL EMBL J. 14:5908-5919(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Marcus S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Squoros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [4]
RP SEQUENCE OF 119-658 FROM N.A.
RX MEDLINE=95320235; PubMed=7597098;
RA Marcus S., Polverino A., Chang E., Robbins D., Cobb M.H.,
RA Wigler M.;
RT "SHK1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian
RT p66PAK protein kinases, is a component of a Ras/Cdc42 signaling
RT module in the fission yeast Schizosaccharomyces pombe.";
RL proc. Natl. Acad. Sci. U.S.A. 92:6180-6184(1995).
CC -!- FUNCTION: FORMS AN ACTIVATED COMPLEX WITH GTP-BOUND RAS-LIKE
CC CDC42. ACTS IN SIGNAL TRANSDUCTION. INVOLVED IN THE CONTROL OF
CC CELL POLARITY AND MATING. MAY INTERACT WITH BYR2.
CC -!- PTM: AUTOPHOSPHORYLATED ON SERINE RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U22371; AAC49125.1; -
CC EMBL: AL034433; CAA22347.1; -
CC EMBL: L41552; AAB52609.1; -
CC HSSP: Q00534; 1B18.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR000095; PAKbox/Rhodog.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00786; PKD; 1.
CC PRINTS: PR00109; TYRKINASE.
CC PRODom: PD0000001; Euk_pkinase; 1.

```


[1]
SEQUENCE FROM N.A.
RC STRAIN-White leghorn; TISSUE=Limb bud;
RA MEDLINE=93300846; PubMed=8314802;
RX Shinomura T., Nishida Y., Ito K., Kinata K.;
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multiforms of PG-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and
CC V1; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60226; CAA42787.1; -;
DR EMBL; D13542; BAA02742.1; -;
DR HSPD; P00740; IEDM.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR ProDom; PD0009918; Link; 2.
DR SMART; SM000032; CCP; 1.
DR SMART; SM000034; CLECT; 1.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM000001; EGF_like; 1.
DR SMART; SM000409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 3562
FT DOMAIN 37 136
FT DOMAIN 166 243
FT DOMAIN 264 345
FT LINK 2

FT DOMAIN 3254 3290 EGF-LIKE 1. (POTENTIAL).
 FT DOMAIN 3292 3328 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3341 3355 C-TYPE LECTIN.
 FT DOMAIN 3460 3518 SUSH1.
 FT DISULFID 44 129 BY SIMILARITY.
 FT DISULFID 171 242 BY SIMILARITY.
 FT DISULFID 195 216 BY SIMILARITY.
 FT DISULFID 269 344 BY SIMILARITY.
 FT DISULFID 293 314 BY SIMILARITY.
 FT DISULFID 328 3278 BY SIMILARITY.
 FT DISULFID 3280 3289 BY SIMILARITY.
 FT DISULFID 3296 3307 BY SIMILARITY.
 FT DISULFID 3301 3316 BY SIMILARITY.
 FT DISULFID 3318 3327 BY SIMILARITY.
 FT DISULFID 3334 3345 BY SIMILARITY.
 FT DISULFID 3362 3454 BY SIMILARITY.
 FT DISULFID 3430 3446 BY SIMILARITY.
 FT DISULFID 3461 3504 BY SIMILARITY.
 FT DISULFID 3490 3517 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 948 948 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 485 1411 MISSING (IN ISOFORM V1).
 SQ SEQUENCE 3562 AA; 388078 MW; 9BC566E880C1602D2 CRC64;
 Query Match 3.4%; Score 106; DB 1; Length 3562;
 Best Local Similarity 22.7%; Pred. No. 41;
 Matches 118; Conservative 61; Mismatches 191; Indels 150; Gaps 29;
 QY 1 MASFTATA-AVSGRWLGNGHNTOPPLSSGSSDLSYCSLPMASRVTRKLVSSALHT--- 56
 DB 1231 MLSSFTAGSILLTLGASPSOTPGSGISSELEEVKIVPFSRAUDKTVTISLDTSSI 1290
 QY 57 -----PPALHFPKSSNSPAIVKPKAKESNTKQMNLI-----PQRAAAALDAAEGL 104
 DB 1291 SAVDKIQPTASKPVSSKSPRII--PEBDEVTSSDIIVIDESISPKASAEEDLTGKM 1348
 QY 105 VSHEKLHLPK-----TADSVQIAGN-FAPVNEQPVRRNLPVVGKLPDS- 148
 DB 1349 VEPE-----IDKBYFTSSTATAVARTAPTPTVMEATEALQPVSPSTSH-----PDSG 1396
 QY 149 --IKGVYVRNGANLPHETVGH-HFFDGDGMVHAFHFGSASACRFQTQTRFVQERQ- 204
 DB 1397 TDIRLVVITGNDTHDPNFEIDLFESRHLPLBANDETHDAESAGTECTSDSVQDSSE 1456
 QY 205 --LGRPVFPKAGELHGHTGIARLMLFYARAAAGIVDPAHGTCVAN-AGLVYFNGRLLAM 261
 DB 1457 YIIIDPFEP-----NFMDFEIEEE-----DCENTTDTVTTPPALQFNGK----- 1495
 QY 262 SEDDLPIYQVITPBGDLKTVGRFEDGQLESTMIAPHKVDPESGELFALSYDVWSKPYLK 321
 DB 1457 YIIIDPFEP-----NFMDFEIEEE-----DCENTTDTVTTPPALQFNGK----- 1495

Db 1496 -----QQVTSAPKSKAEARSD---QIES--VAHSK-NVTFSQLNTEFTIISFT--- 1540
 QY 322 YFRFSPDGTGKSPDV-----ETQLDQPTMMHDFAITENFVVVVDQOVVKPLPMMINGGSP 375
 Db 1541 ----EASGTMQPSKAGVMGAFEVTPQTA--DVAMLE-----P 1572
 QY 376 VYDKNKVARFGILDKYAEEDSS-----NIKWDAPDFCFHLLWANEE---PETDEV 424
 Db 1573 VYSGEVEI---TTDKYLEITSVYEQSPKKNKTV-----MMHGTESSYTKTKNKL 1620
 QY 425 VVIGSCMTPPDSIFNESDENLKSILSEIRLNKLTGSTRR 464
 Db 1621 LLITNESSGDGS--TESDLS-RSVFTEI-LTMSSHEDSEK 1656
 RESULT 6
 CAD5_MOUSE
 ID CAD5_MOUSE STANDARD; PRT; 783 AA.
 AC P55284; O35542;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Vascular endothelial-cadherin precursor (VE-cadherin) (Cadherin-5).
 OS CDH5.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain capillary;
 RX MEDLINE=96141083; PubMed=8555485;
 RA Breier G., Breviaro F., Caveda L., Berthier R., Schnuerch H.,
 RA Gotsch U., Vestweber D., Risau W., Dejana E.;
 RT "Molecular cloning and expression of murine vascular endothelial-
 RT cadherin in early stage development of cardiovascular system.";
 RL Blood 87:630-641(1996).
 RN [2]
 RP SEQUENCE FROM N.A. AND FUNCTION.
 RC STRAIN=BA16/c; TISSUE=Breast carcinoma;
 RX MEDLINE=97364256; PubMed=9220534;
 RA Matsuyoshi N., Toda K.-I., Horiguchi Y., Tanaka T., Nakagawa S.,
 RA Takeichi M., Imamura S.;
 RT "In vivo evidence of the critical role of cadherin-5 in murine
 RT vascular integrity";
 RL Proc. Assoc. Am. Physicians 109:362-371(1997).
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
 CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS.
 CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
 CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X83930; CAA58782.1; -
 CC EMBL: D63942; BAA22617.1; -
 CC HSSP: P15116; INCU.
 CC MGD: MGI:105057; Cdh5.
 CC InterPro: IPR002126; Cadherin.
 CC InterPro: IPR000233; Cadherin_C_term.
 CC Pfam: PF00028; cadherin; 5.
 CC Pfam: PF01049; Cadherin_C_term; 1.
 CC PRINTS: PR00205; CADHERIN.
 CC SMART: SM00112; CA; 5.

DR	PROSITE; PS00232; CADHERIN_1; 3.
DR	PROSITE; PS50268; CADHERIN_2; 5.
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW	Signal.
FT	SIGNAL 1 24 POTENTIAL.
FT	PROPEP 25 45 POTENTIAL.
FT	CHAIN 46 783 VASCULAR ENDOTHELIAL-CADHERIN.
FT	DOMAIN 46 592 EXTRACELLULAR (POTENTIAL).
FT	TRANSEM 593 619 POTENTIAL.
FT	DOMAIN 620 783 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 620 783 CADHERIN 1.
FT	DOMAIN 149 255 CADHERIN 2.
FT	DOMAIN 256 370 CADHERIN 3.
FT	DOMAIN 371 475 CADHERIN 4.
FT	DOMAIN 476 592 CADHERIN 5.
FT	DOMAIN 737 752 SER-RICH.
FT	CARBOHYD 59 59 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 154 154 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 440 440 CARBOHYD (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT 67 69 KDO -> GKX (IN REF 2).
SEQ	SEQUENCE 783 AA; 87847 MW; 240AF2D663BCE71C CRC64;
Query Match	
Best Local Similarity 13.3%; Score 105; DB 1; Length 783;	
Matches 128; Conservative 101; Mismatches 242; Indels 254; Gaps 32;	
QY	60 LHPFKQSSNPAIVKPKAKESNTKOMNLFORAAA---ALDAAGFLVSHKLHP----- 112
DB	52 MHIDEENESLPVHKQSNVRONAKYVLGEFAGKIGVDANGNVLAVERLDREKVS 111
QY	113 -----LPKTADPSVO-----IAGNFAPVNEQPVRRNLPVVGKLPDSIKGY 153
DB	112 EYFLTALIVDKNTKNLEQPSSFVKVHDINDNPVFESHQVFNASVPMSAIGSVIRVT 171
QY	154 VRNGANP-----LHEPVTGHFF--DGCMHWAKFEHGSASYACRFTQTNRFOER 203
DB	172 AVDADDPTVAGHAVLVQIVRGNEYFSIDNSGLI-----FTKIKNLDEK 216
QY	204 QLGRPFVPFK---AIGELHGHTGIARLMFLYARAAGAIVDPAHGTGVANAGLVYFNGLLA 260
DB	217 QAELYKIIVETDALG-LRGESGTAIVMI-----RL-- 245
QY	261 MSEDDLFPYOVIITNGOLKTVRGFDFGQLBSTMLAHP-----KVDPSEGFALSYDVV 315
DB	246 --ED-----INDNFPVFTOSTYTFSP-EDIARGKPLGLFTVVDPDPQNRTKYSIM 295
QY	316 SKPLYKFREFSPD-----GTKSPDVE-----IQLDQPTMHMF----- 348
DB	296 QGEYRDFTETIDPKRNEGIIKPKSLDYEQYTFVIEATDTPIREYLSSTSGKKA 355
QY	349 AITENFVVVDPOVV-----FKLPEKIRG---GSPVVYDNKNVAR-FGLDKYAEDSS- 397
DB	356 MVTINVLDVDDPPPYQRHFYHFKLPENQKKPLIGTVAAKDPKQSRISYIRKTSDRGQ 415
QY	398 -----NIKWDAPDCFCFHNNAWEEPETE-----VVIGSCMT 432
DB	416 FFRITKGNIYNKELDREITYAVNL--TVBANELDSRGNPVGKESIVQVYIEVLNDEN 473
QY	433 PPDSI-----FNESDENLKSVLSIRLNLTKGESTREPIIISNQDQ 473
DB	474 PPEFAQYPVKVCENAAGQKLVQISAIDKDQVVPNPKFKALK-NEDSNTFLNNHNT 532
QY	474 VNL--EAGWNRNMLGRKTKFAYIALA---BPWKVSGFAKVDLTGTGVKKHLYGDNRYG 528
DB	533 ANITVKYQGPNR---EHAKPHYLPLVISDGNVPSLTGTS--TLTVGVCKNEQGEFTFC 586
QY	529 GEPLFLPGEGGEDGYILCFV-----HDE 553
DB	587 EEMAQAQGVSIQALVAIFLCILTITVTLLILRRRIKQHAHSKSALETHEGLVLYDE 646
QY	554 K-----TWKSELQIVNAV-----SLEVEATVKLPSRPVPGFHGTFIGADD 593

```

CC DR EMBL: U26644; AAC50259.1;
DR Genew: HGNC:3594; RASN.
DR MIM: 600212;
DR InterPro: IPR001227; Ac-transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR003880; Pantine_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; pp-binding; 1.
DR Pfam: PF00698; Acyl_transfer; 1.
DR Pfam: PF00975; Thioesterase; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR PROSITE: PS00606; B-KETOACYL-SYNTHASE; 1.
DR PROSITE: PS00075; ACP_DOMAIN; 1.
DR Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW Pyridoxal phosphate.
FT DOMAIN 1 413
FT DOMAIN 428 815
FT DOMAIN 1630 1857
FT DOMAIN 1858 2113
FT DOMAIN 2118 2174
FT DOMAIN 2202 2504
FT ACT_SITE 161
FT ACT_SITE 580
FT ACT_SITE 876
FT NP_BIND 1666 1683
FT BINDING 1699
FT NP_BIND 1880 1895
FT BINDING 2151
FT ACT_SITE 2302
FT ACT_SITE 2475
FT SEQUENCE 2504 AA; 273100 MW; 8AAF9809B2338DFA CRC64;

Query Match
Best Local Similarity 3.3%; Score 105; DB 1; Length 2504;
Matches 81; Conservative 41; Mismatches 143; Indels 82; Gaps 19;

QY 11 SGRWLGNNHTQPTLLSSQSSDLSYCSLPLMASRVTKLN--VSSALHTPPLHFPKQS-- 66
Db 707 SARWL-----STSPDAQWSSSLARTSSAEYNNVNLSPVLFQFALWVPEHAVV 756
QY 67 -----SNSPAIVVVKAKESNTKQMNLFQRAAAALDAAGFLVSHEKHLPLKPTADPSV 121
Db 757 LEIAPTQCPQAVLKVRKPSCT-----IIPRMKKDHRDNEFFLAGTGRHLHSLCIDANPVA 812
QY 122 QIAGNFAPVNEOPVRNLPVVGKLPDSIKGVVVRGANGFLHEPVTGHHFFDGDGMVHAVK 181
Db 813 L-----FPPV-ESPARGTPLISPL-----IKWDHSLWADAPA-AED 847
QY 182 FHGSGAS-----YAC--RFTQNRFRVQERQL--GRVPFKPAIGELHGHGTGIARMLFYARA 233
Db 848 FPNQSGSPSATIYCTPSSSDPRIVNDITDGRVLPAT-----GYLSIVWKTLL--ARA 900
QY 234 AGI-----VDPAGTGVANAGLVYNGRLAMSEDDLPYQVQITPNGDGLKTVGR- 283
Db 901 NAGLEQLFVVFDDVQVHQHQTILPKTGTVSLEVRLL-----EATGAFSENGNLVVGKV 955
QY 284 FDFDQGLESTMTAKPK-VDPES--GELFALSVDVYKPKYLFERFSP 327
Db 956 YQWDDP-DPRFDHPSPHPNPSRSPFLFQAQAEVYKELRLRGDYGP 1001

RESULT 8
ID APCE_FREDI
AC P16566;
DT 01-AUG-1990 (Rel. 15, Created)

```

```

DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 01-DEC-1992 (Rel. 24, Last annotation update)
DE Phycobillosome 120 kDa linker polypeptide, core (L-CM 92) (Core-
DE membrane linker protein).
GN APCE.
OS Fremyella diplosiphon (Calothrix PCC 7601).
OC Bacteria: Cyanobacteria; Nostocales; Rivulariaceae; Fremyella.
OX NCBI_TaxID=1197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192765; PubMed=2107546;
RA Houtard J., Capuano V., Colombano M.V., Coursin T., de Marsac N.;
RT "Molecular characterization of the terminal energy acceptor of
RT cyanobacterial phycobillosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2152-2156(1990).
CC -1- FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL
CC ENERGY ACCEPTOR (BY ITS PHYCOBILIN-LIKE DOMAINS) AND AS A LINKER
CC POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE
CC PHYCOBILISOME CORE ARCHITECTURE.
CC -1- SUBUNIT: PHYCOBILISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO
CC CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY
CC COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS, AND OF THREE
CC MINOR COMPONENTS: THE ALLOPHYCOCYANIN ALPHA-B CHAIN, A 18.3 kDa
CC POLYPEPTIDE, AND THE ANCHOR POLYPEPTIDE LCM.
CC -1- SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILISOME PERPENDICULARLY
CC TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC -1- SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL
CC REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDES.
CC -1- SIMILARITY: THE PHYCOBILIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILINS
CC FROM VARIOUS SPECIES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M20806; AAA24873.1;
CC PIR: A35088; A35088.
CC HSP: P00318; 1B33.
CC InterPro: IPR001297; PBS_linker_poly.
CC InterPro: IPR001659; Phycobillosome.
CC Pfam: PF00427; PBS_linker_poly; 4.
CC Pfam: PF00502; Phycobillosome; 1.
CC ProDom: P000340; Phycobillosome; 2.
CC Phycobillosome; Electron transport; Photosynthesis; Repeat.
KW INIT_MET 0 0
FT DOMAIN 17 75 PHYCOBILIN-LIKE 1.
FT DOMAIN 76 143 PHYCOBILIN-LIKE LOOP.
FT DOMAIN 144 236 PHYCOBILIN-LIKE 2.
FT DOMAIN 237 284 ARM 1 (SPACING SEQUENCE).
FT REPEAT 285 409 I.
FT DOMAIN 410 546 ARM 2 (SPACING SEQUENCE).
FT REPEAT 547 669 II.
FT DOMAIN 670 743 ARM 3 (SPACING SEQUENCE).
FT REPEAT 744 869 III.
FT DOMAIN 870 953 ARM 4 (SPACING SEQUENCE).
FT REPEAT 954 1079 IV.
FT SEQUENCE 1079 AA; 120325 MW; 87FE38F232BF82 CRC64;

Query Match
Best Local Similarity 3.3%; Score 104.5; DB 1; Length 1079;
Matches 71; Conservative 47; Mismatches 80; Indels 135; Gaps 21;

QY 62 FPKQSSN---SPAIVVVKPAKESNTKQMNLFQRAAAALDAAGFLVSHEKHLHP-----L 113
Db 467 FPKETRNFTSPA---PFSKDRRLIN-----OGPGINSQVSNPGARGE 509
QY 114 PKTADPSV-----QIAG-----NFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNG 157
Db 510 PGSGLPKVFRDLQPLGTIGKKAAGKASIKPSESSTQAV-----IKAYLQ-- 554

```

QY	158	ANLPEVTVGHFFDGDGM-VHNVFPEHGSAS-----YACRET	194
DB	555	-----VFGRVYEGQRLKVOEIKLENGQLSVREFIRALAKSDVERKYWTSLSYVCKAI	607
QY	195	QTNRFVQERQLGRPVFPEKAIHELHGHGTGLARLMLFYARAAAGIVDPAHGTGVANAGLVYF	254
DB	608	E---YTHRLLRPTVGRQ--EINKYFDAAKOGFYA-----VVD-----AIIINS--VEY	650
QY	255	NGRLAMSEDDLPYQVOITPNG-----DLKTVG--RDFDQGLESTMIAPKVPDPESGELF	308
DB	651	SE---AFGEDTVPYRYLTGPGVALRQLR-VGSIREDVGS-----KVOKQETPLF	696
QY	309	ALSYDVVSRPKLYKFRFSPDKSPDVEIQLDQ	341
DB	697	VTLGTVT-----DTRTPDQIFRINQ	717
RESULT 9			
ID	SYFB_AQUAE	STANDARD;	PRT; 775 AA.
AC	O67620;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--		
DE	tRNA ligase beta chain) (PheRS).		
GN	PHE1 OR AQ_1730.		
OS	Aquifex aeolicus.		
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;		
OC	Aquifex.		
OX	NCBI_TaxID=63363;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-VF5;		
RA	MEDLINE=98196666; PubMed=9537320;		
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,		
RA	Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,		
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;		
RT	"the complete genome of the hyperthermophilic bacterium Aquifex		
RT	aeolicus."		
RT	Nature 392:353-358(1998).		
CC	-!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +		
CC	diphosphate + L-phenylalanyl-tRNA(Phe).		
CC	-!- SUBUNIT: Tetramer of two alpha and two beta chains (BY		
CC	SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA		
CC	CHAIN FAMILY. SUBFAMILY 1.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF000753; AAC07582.1;		
DR	HSSP; P27002; 1PYS.		
DR	InterPro; IPR005146; B3_4.		
DR	InterPro; IPR005147; B5.		
DR	InterPro; IPR005121; Fdx-Anticb.		
DR	InterPro; IPR004532; PheT_Bact.		
DR	InterPro; IPR002547; tRNA_bind.		
DR	Pfam; PF01588; tRNA_bind; 1.		
DR	Pfam; PF03147; FDX-ACB; 1.		
DR	Pfam; PF03483; B3_4; 1.		
DR	Pfam; PF03484; B5; 1.		
DR	TIGRfam; TIGR00472; pheT_bact; 2.		
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; AIP-binding.		
KW	Complete proteome.		
QY	SEQUENCE 775 AA; 88484 MW; 61BB15FDE5B274C8 CRC64;		

Query Match

3.3%; Score 103; DB 1; Length 775;

Best Local Similarity

20.7%; Pred. No. 7.4;

Matches 114; Conservative 67; Mismatches 159; Indels 212; Gaps 31;

QY	123	IAGNAPVNEQPVRRNLPVVGLPDSIKGVYVRNGANPLHEPVTHGHFFD-----	172
DB	46	VFGKVVEVKEHTPKKLAIV-----KVQVEHIFIDVVTVDKSVRE	86
QY	173	GDGMVHAVFEHGSASACRETQTNRFVQERQLGRPVFPEKAIHELHGHGTGLARLMLFYAR	232
DB	87	GDGVI--VALPNKAVGNMC-----VTEREF-----DGVSCKGLLSA	121
QY	233	AAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLP-----YQVOITN-GD	277
DB	122	QELGLEEKSEG-----VLKIHEDFKPGTDANEILGFGKEKIIEDITPNRGD	167
QY	278	LKTVGREFDQGLESTMIAPKVPDP-----ESGELFALSYDVVSRPKLYKFRFSPDKGK	331
DB	168	MLSVRGVARD---LSAIFRLPKKKPEPTYEETGETFEIEDCKKY-----RGVV	216
QY	332	SPDVEIQLDQP-----TMMHDFAITENFVVVVDQVVFVKLPPEMIRGSPV-VY	378
DB	217	IEGVEIK-ESPLYKKRLWQGGIKSINNVDITNYV-----MLRGQPLHAF	262
QY	379	DNKNVAREGILDYAESSNKKWIDAPDCFCFHLNNAWEPEDEVVVIGSCMTPPDSIF	438
DB	263	DLSKV-EGGIIVRSAKKGEKIITLDG-----EERELEDILVITADREKPLAVA	309
QY	439	N-----ES--DENLKSVLSEI-----RLNLKTGSTRPRPIISNEDQOVNLEA	478
DB	310	GVIGGLESGKENTKIDILLESAYENPFRVRKASKKLGIOLESSYR-----FERNVDIE-	362
QY	479	GMVNRNMLGRKTKFA-YIALAEPKPVSGPAKVDLTGTGEVKKHLYGDNRYGGEPLFLP--	535
DB	363	-----RVDRADQYAVVILKHKAGKV-----KVKQVYRE-KYKPKKVELPQG	404
QY	536	-----GEGGEDEGYILCFVDEKTKWSELQIVNAVSLVEATVKLPSRVPYGFHGTFF-	588
DB	405	KYIRYAGESYKNEE-----VKREILDALIEIPNEI-MRCGVEVLVPS-----HRSFD	448
QY	589	IGAD-DIAKQV 599	
DB	449	IQRVDLIEEM 460	

RESULT 10

CYAA_USTMA

STANDARD;

PRT; 2493 AA.

ID	CYAA_USTMA	STANDARD;	PRT; 2493 AA.
AC	P49606;		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl		
DE	cyclase).		
GN	UAC1 OR REM1		
OS	Ustilago maydis (Smut fungus).		
OC	Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;		
OC	Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.		
OX	NCBI_TaxID=5270;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-518;		
RX	MEDLINE=95087882; PubMed=7995519;		
RA	Gold S., Duncan G., Barrett K., Kronstad J.W.;		
RT	"cAMP regulates morphogenesis in the fungal pathogen Ustilago		
RT	maydis."		
RL	Genes Dev. 8:2805-2816(1994).		
CC	-!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR		
CC	METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,		
CC	CAMP.		
CC	-!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.		
CC	-!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.		

```
CC -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC -!- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L33918; AAA57469.1; .
DR InterPro; IPR001054; G_Cyclase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR001932; PP2C-like.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00211; guanylate_cyc; 1.
DR Pfam; PF00481; pp2c; 1.
DR Pfam; PF00560; LRR; 17.
DR SMART; SM000044; CYCc; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00369; LRR_TYP; 2.
DR SMART; SM00332; PP2C; 1.
DR SMART; SM00314; RA; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
FT DOMAIN 759 763 POLY-ASP.
FT REPEAT 888 897 POLY-ALA.
FT REPEAT 1084 1106 LRR 1.
FT REPEAT 1108 1131 LRR 2.
FT REPEAT 1132 1155 LRR 3.
FT REPEAT 1157 1178 LRR 4.
FT REPEAT 1179 1201 LRR 5.
FT REPEAT 1202 1225 LRR 6.
FT REPEAT 1227 1248 LRR 7.
FT REPEAT 1249 1271 LRR 8.
FT REPEAT 1273 1295 LRR 9.
FT REPEAT 1314 1337 LRR 10.
FT REPEAT 1338 1360 LRR 11.
FT REPEAT 1361 1384 LRR 12.
FT REPEAT 1386 1407 LRR 13.
FT REPEAT 1408 1430 LRR 14.
FT REPEAT 1432 1455 LRR 15.
FT REPEAT 1509 1529 LRR 16.
FT REPEAT 1533 1556 LRR 17.
FT REPEAT 1557 1580 LRR 18.
FT REPEAT 1581 1604 LRR 19.
FT REPEAT 1606 1628 LRR 20.
FT REPEAT 1633 1656 LRR 21.
FT REPEAT 1722 2001 PP2C-LIKE.
FT DOMAIN 2002 2493 CATALYTIC.
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5B8F8 CRC64;

Query Match          3.2%; Score 102; DB 1; Length 2493;
Best Local Similarity 21.7%; Pred No. 48;
Matches      85; Conservative 44; Mismatches 136; Indels 126; Gaps 17;

QY    4   FTTAAVSGRWLGNNIT-----QPPLSSQSDLSYCSSLPMASRVTRKLNVSALHTP 57
Db    801 FFKTASLAGSRRTDUSDVDLTALPLPGSKSVDEAAANKVDVLQQ-TNNAQSALVQQ- 858
QY    58 PALHFPKSSNSPAIVVKPKAESNTIKMNLQRAAAAALDAAEGLVSH-----SKLHPL 113
Db    859 ---QSOSQHHPQSPNVRTTSRGGAHM--FASAGASAAAAGKLGLHRPSKRMNAR 913
QY    114 PKTAD-----PSVQIAGNFAPVNEQPVRRNLVVVGKLPDSIKGYVVRGANGFLHEPTG 167
Db    914 PNTAGSVGATRPSTTLTGSLTAEDD-----TSINGSTRRDG-HPLKRDATA 959
QY    168 -----HHFFDGDMGVIAVKEHGSAASYACRFTQINRFVQGRQLGRVPFKAGE 216
```


[illegible]

DR PIR; A40889; A40889.
 DR HSP; P14687; 1AMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR003880; Pantine_attach.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR InterPro; IPR001031; thioesterase.
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00658; Condensation; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 DR PROSITE; PS00455; AMP_BINDING; 3.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.
 KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
 KW Repeat; Phosphopantetheine.
 FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
 FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
 FT DOMAIN 950 919 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
 FT BINDING 882 882 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1965 1965 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3050 3050 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT ACT_SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3770 AA; 422448 MW; CB66B6D232A58CB0 CRC64;

Query Match 3.2%; Score 100; DB 1; Length 3770;
 Best Local Similarity 20.9%; Pred. No. 1.2e+02;
 Matches 97; Conservative 67; Mismatches 162; Indels 138; Gaps 27;

Qy 12 GRWLGNNHTQPLSSSSDLSYCSLLPMASRVTRK-----INVSAHLHPAL 60
 Db 1544 GNWSVSN-----PAPSSTSLDAY-----IYTSOTGKPKGVVHGVVNVNLOISLSKIFGL 1596
 Qy 61 -----HFPKQSS-----NSPAIVVKKPKAKESNTKQMNLFORAAAAALDA 99
 Db 1597 RDTDEVILSFNSVYDFHFVQMTDAILNGQTLVNLDMRSKRER--LYQYIETNRVY 1654
 Qy 100 AGEFLVSHEKHLPLFKTADPSQVIAGNFAPVNEQPVRR-----NLFVVGKLPDSIKG 151
 Db 1655 LSG-----TPSVISMYEFSREKDH--LRVDCVGEAFSQPVDFQIDRTQG 1698
 Qy 152 VYVRNCAANPLHEPVYTGHHFFDGDGMVHAKFEHSASVACRFTQNRNRFVQERQLOE-PVF 210
 Db 1699 LII-NGYGTPEISITHK-----RLYPPERRTDKSIQGIQNSTSVYLNADMKRVPI- 1750
 Qy 211 PKAIGELH-GHTGIARLMFYAR-----AAAGIVDPAGHTGVANAGLVYFNGRLLAISED 265
 Db 1751 -GAVGELYLGEGVAR--GYHNRPEVTAERLRNPFQIDSEQRNGR---NSRLYETGD-- 1802
 Qy 266 LPYQVOIITP--NGDLKTVGRDFD-----GQLESTMIAPH-----KVDPE 304
 Db 1803 ---LVRVTPGSGNETEYLGRLNDFQVIRGLRIELGEIEAVMSHSDPIKQSVVIAKSGREG 1859
 Qy 305 GELFALSVDVYVSKPKLYKFRSPDGTKSPDVFIQLDQDTMMHDFALTENFV-----VVP 358
 Db 1860 DQKFLVGEVASSP-----LSPGA-----IRFMOSRLPGYMPISPISSLPVTP 1906
 Qy 359 DOOVYFK-LPEM-IRGGSFVYDKNKVAF-----GILDKYAE 394
 Db 1907 SKLDTKALPTAEKGMNVLAAPRNEIESILGIGISAGLIDISAQ 1950

RESULT 15
 ID PHYC_BACSU STANDARD; PRT; 383 AA.
 AC Q31097;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

3-phosphate precursor (EC 3.1.3.8) (Phytate 3-phosphatase) (MYO-
 inositol-hexaphosphate 3-phosphohydrolase).
 PHYC OR PHYB13.
 Bacillus subtilis.
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=1423;
 (1)
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 STRAIN=VTT-E-68013;
 MEDLINE=98268943; PubMed=9603817;
 Kerovuo J., Lauraeus M., Nurminen P., Kalkkinen N., Apajalahti J.;
 "Isolation, characterization, molecular gene cloning, and sequencing
 of a novel phytase from *Bacillus subtilis*."
 Appl. Environ. Microbiol. 64:2079-2085(1998).
 CC -!- FUNCTION: CATALYZES THE HYDROLYSIS OF INORGANIC ORTHOPHOSPHATE
 FROM PHYTATE. ONLY PHYTATE, ADP, AND ATP WERE HYDROLYZED (100, 75,
 AND 50% OF THE RELATIVE ACTIVITY, RESPECTIVELY). HAS MAXIMAL
 ACTIVITY AT PH 7 AND 55 DEGREES CELSIUS.
 CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = D-myo-
 inositol 1,2,4,5,6-pentakisphosphate + phosphate.
 CC -!- COFACTOR: ISOLATED ENZYME REQUIRED CALCIUM FOR ITS ACTIVITY AND/OR
 STABILITY.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- INDUCTION: BY PHYTATE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AF029053; AAC31775.1; -
 HSSP; O66037; 2POO.
 InterPro; IPR003431; Phytase.
 Pfam; PF02333; Phytase; 1.
 Hydrolase; Signal.
 SIGNAL 1 26 POTENTIAL.
 PROPEP 27 29
 FT CHAIN 30 383 3-PHYTASE.
 SQ SEQUENCE 383 AA; 41922 MW; E9BEC2E4A48EB9CA CRC64;

Query Match 3.2%; Score 99.5; DB 1; Length 383;
 Best Local Similarity 19.3%; Pred. No. 4.8;
 Matches 88; Conservative 71; Mismatches 151; Indels 147; Gaps 22;

Qy 18 NHTQPLSSSSDLSYCSLLPMASRVTRKLN-----VSSALHTPPALHFPKSSNSPA 71
 Db 2 NHSKTLTLLTAAGLMLTCA--VSSQAKHKLSDPVHFTVNAAEETP-VDTAGDAADDP 58
 Qy 72 IVVKKAKESNTKQMNLFORAAAAALDAEGLVSHKHLPLPKTADPSVQIAGNFAPVN 131
 Db 59 IWLDEKTPQ-NSK-----LIITNKKSLGVVYSLDGKMLHSYNTGKLN 99
 Qy 132 EQPVERNPPVVGKLPDSIKGVVVRNCAANPLHEPVYTGHHFFDGDGMVHAKFEHSASVAC 191
 Db 100 NVDIRVDFPLNGKVKYDIAAASNRSEKNTIE-----IYALDGKNGTL--- 141
 Qy 192 RETQTNRFVQERQLOEGRVFP--KAIGELHGTGIARLMFYARAAAGIVDPAGHTGVANA 249
 Db 142 -----QSMTPDHPHATATATINEVYGT-----LYHSQK----- 168
 Qy 250 GLIVFNGRLAM---SEDDL-PYOVQIITPNDL--KTGREFDQGOLESTMIAPHKVDPE 303
 Db 169 -----TGKIYAMVTGKEGEFEQYELKADKNGYISGKKVRAFNMNSQTEGM-----AAODE 218
 Qy 304 SGELFALSVDVYVSKPKLYKFRSPDGTKSPDVFIQLDQDTMMHDF-----AITENFV 355
 Db 219 YGRLYIAFEDEA-----IWKFSAEPPGGSGNGTVIDRADRGHLTRDIEGLTIYVADGKYL 274
 Qy 356 VVPDQOVVKLPFEMIRGGSPVYD---KNK-VAREGILDKVAEDSSNIKWIDAPDPCFCH 411

Db 275 MASSQ-----GNSSAIYDRQGNKYVAFRITDGPETDGT----- 311

QY 412 LWNABEPETDEVVIGSCMTP--PDSIFNED-ENL 445

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	35.2	12.5	22	65
Gender	0.48	0.50	0	1
Marital Status	0.65	0.48	0	1
Education	12.8	2.1	9	16
Income	45000	15000	20000	80000
Health	0.72	0.45	0	1
Smoking	0.25	0.43	0	1
Alcohol	0.18	0.38	0	1
Exercise	0.35	0.48	0	1
Stress	0.60	0.49	0	1
Depression	0.22	0.42	0	1
Loneliness	0.30	0.46	0	1
Life Satisfaction	0.55	0.50	0	1
Quality of Life	0.68	0.47	0	1
Overall Health	0.75	0.44	0	1
Physical Health	0.80	0.40	0	1
Mental Health	0.60	0.49	0	1
Social Health	0.50	0.50	0	1
Emotional Health	0.45	0.51	0	1
Psychological Health	0.40	0.52	0	1
Behavioral Health	0.35	0.48	0	1
Environmental Health	0.30	0.46	0	1
Community Health	0.25	0.43	0	1
Natural Health	0.20	0.40	0	1
Artistic Health	0.15	0.35	0	1
Spiritual Health	0.10	0.30	0	1
Philosophical Health	0.05	0.25	0	1
Scientific Health	0.02	0.15	0	1
Technological Health	0.01	0.10	0	1
Medical Health	0.00	0.05	0	1
Legal Health	0.00	0.05	0	1
Political Health	0.00	0.05	0	1
Economic Health	0.00	0.05	0	1
Cultural Health	0.00	0.05	0	1
Religious Health	0.00	0.05	0	1
Historical Health	0.00	0.05	0	1
Geographical Health	0.00	0.05	0	1
Biological Health	0.00	0.05	0	1
Chemical Health	0.00	0.05	0	1
Physical Health	0.00	0.05	0	1
Mental Health	0.00	0.05	0	1
Social Health	0.00	0.05	0	1
Emotional Health	0.00	0.05	0	1
Psychological Health	0.00	0.05	0	1
Behavioral Health	0.00	0.05	0	1
Environmental Health	0.00	0.05	0	1
Community Health	0.00	0.05	0	1
Natural Health	0.00	0.05	0	1
Artistic Health	0.00	0.05	0	1
Spiritual Health	0.00	0.05	0	1
Philosophical Health	0.00	0.05	0	1
Scientific Health	0.00	0.05	0	1
Technological Health	0.00	0.05	0	1
Medical Health	0.00	0.05	0	1
Legal Health	0.00	0.05	0	1
Political Health	0.00	0.05	0	1
Economic Health	0.00	0.05	0	1
Cultural Health	0.00	0.05	0	1
Religious Health	0.00	0.05	0	1
Historical Health	0.00	0.05	0	1
Geographical Health	0.00	0.05	0	1
Biological Health	0.00	0.05	0	1
Chemical Health	0.00	0.05	0	1
Physical Health	0.00	0.05	0	1
Mental Health	0.00	0.05	0	1
Social Health	0.00	0.05	0	1
Emotional Health	0.00	0.05	0	1
Psychological Health	0.00	0.05	0	1
Behavioral Health	0.00	0.05	0	1
Environmental Health	0.00	0.05	0	1
Community Health	0.00	0.05	0	1
Natural Health	0.00	0.05	0	1
Artistic Health	0.00	0.05	0	1
Spiritual Health	0.00	0.05	0	1
Philosophical Health	0.00	0.05	0	1
Scientific Health	0.00	0.05	0	1
Technological Health	0.00	0.05	0	1
Medical Health	0.00	0.05	0	1
Legal Health	0.00	0.05	0	1
Political Health	0.00	0.05	0	1
Economic Health	0.00	0.05	0	1
Cultural Health	0.00			

```

Db 312 -----DTDGDVLGFLGPEYFFGIFVAQDGENI 340

```

Search completed: May 19, 2003, 15:48:01
Job time : 45 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 15:46:14 ; Search time 43 Seconds
(without alignments)
2870.284 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATATVSGRWLGGNHT.....VPGFHGTFIGADDLAKQV 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150	100.0	599	Q9LRR7	Q9LRR7 arabidopsis
2	3143	99.8	599	Q93ZU5	Q93ZU5 arabidopsis
3	2289	72.7	604	Q9M329	Q9M329 solanum tub
4	2280.5	72.4	605	Q24023	Q24023 lycopersico
5	2170.5	68.9	657	Q9M9F5	Q9M9F5 arabidopsis
6	2168.5	68.8	612	Q9FS24	Q9FS24 vigna ungui
7	2165	68.7	615	Q9M6E8	Q9M6E8 phaseolus v
8	2066.5	65.6	625	Q9AXZ4	Q9AXZ4 persea amer
9	2044.5	64.9	589	Q9C6Z1	Q9C6Z1 arabidopsis
10	1991	63.2	583	Q49505	Q49505 arabidopsis
11	1977	62.8	569	Q9AXZ3	Q9AXZ3 persea amer
12	1930	61.3	604	Q24592	Q24592 zea mays (m
13	1663.5	52.8	577	Q9LRR7	Q9LRR7 arabidopsis
14	968	30.7	595	Q49675	Q49675 arabidopsis
15	959.5	30.5	543	Q94IR2	Q94IR2 phaseolus v
16	938	29.8	538	Q9LY63	Q9LY63 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9LRR7	PRELIMINARY;	PRT;	599 AA.
AC	Q9LRR7;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	9-cis-epoxycarotenoid dioxygenase (Neoxanthin cleavage enzyme).			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_taxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-COLUMBIA;			
RA	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-COLUMBIA;			
RX	MEDLINE=20277480; PubMed=10819329;			
RA	Nakamura Y.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence			
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC			
RL	DNA Res. 7:131-135(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-COL;			
RA	Iuchi S., Kobayashi M., Shinozaki K.;			
RT	"Characterization of neoxanthin cleavage enzyme from Arabidopsis			
RT	thaliana."			
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; AB028617; BAB01336.1; -			
DR	EMBL; AB026549; BAB70609.1; -			
DR	InterPro; IPR004294; RPB65.			
DR	Pfam; PF03055; RPB65; 1.			
KW	Dioxygenase.			

065572 arabidopsis
Q94en8 lactuca sat
Q9axz5 persea amer
Q8vxp3 anabaena sp
Q45895 malus domes
Q9aa32 caulobacter
Q9amil streptomyce
Q93fa4 streptomyce
O06785 mycobacteri
Q9K46 streptomyce
O05905 mycobacteri
P74334 synecocyst
Q8vxn9 crocus sati
Q8ypb4 anabaena sp
Q53353 pseudomonas
Q8vxp1 crocus sati
Q8vxp2 arabidopsis
Q8xta2 ralstonia s
P74370 synecocyst
Q8rqw2 pseudomonas
Q52008 pseudomonas
Q93vd5 oryza sativ
O68868 synecococc
Q91993 gallus gall
Q90wn4 brachydanio
Q9m079 arabidopsis
Q9shd9 arabidopsis
Q8vxp0 crocus sati
Q8ymn8 anabaena sp

538 10 065572
200 10 Q94en8
524 10 Q9axz5
475 16 Q8vxp3
446 10 Q45895
483 16 Q9aa32
456 2 Q9amil
456 2 Q93fa4
501 16 O06785
503 16 Q9K46
502 16 O05905
490 16 P74334
209 10 Q8vxn9
497 16 Q8ypb4
485 2 Q53353
218 10 Q8vxp1
570 10 Q8vxp2
520 16 Q8xta2
480 16 P74370
491 2 Q8rqw2
490 2 Q52008
552 10 Q93vd5
342 2 O68868
526 13 Q91993
516 13 Q90wn4
616 10 Q9m079
618 10 Q9shd9
185 10 Q8vxp0
472 16 Q8ymn8

937 29.7
837.5 26.6
798.5 25.3
745.5 23.7
745 23.7
528.5 16.8
527.5 16.7
526.5 16.7
466 14.8
464 14.7
454.5 14.4
394 12.5
382.5 12.1
355.5 11.3
352 11.2
327 10.4
322.5 10.2
319 10.1
317 10.1
311.5 9.9
303.5 9.6
302.5 9.6
270 8.6
260 8.3
256 8.1
253 8.0
244 7.7
242 7.7
242 7.7

SQ SEQUENCE 599 AA; 65856 MW; 7D513F39945E0CF3 CRC64;

Query Match 100.0%; Score 3150; DB 10; Length 599;
 Best Local Similarity 100.0%; Pred. No. 1.8e-237;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAAYSGRWLGNHTQPPPLSSQSSDLSYCSSLPMAASRVTRKLNYSALHTPPAL 60
 DB 1 MASFTATAAAYSGRWLGNHTQPPPLSSQSSDLSYCSSLPMAASRVTRKLNYSALHTPPAL 60

QY 61 HFPKQSSNSPAIVKPKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120
 DB 61 HFPKQSSNSPAIVKPKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120

QY 121 VQIAGNFAFVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VQIAGNFAFVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASACRTQTNRFRVQERQLGRPVPPKAIGELHGTGIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASACRTQTNRFRVQERQLGRPVPPKAIGELHGTGIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGRLAMSEDDLLPYQVQITPNDGLKTGVGRFDGQLESTMIAPKV 300
 DB 241 AHGTGVANAGLVYFNGRLAMSEDDLLPYQVQITPNDGLKTGVGRFDGQLESTMIAPKV 300

QY 301 DPESGELFALSYDVSKPYLYKFRFSPDGTGKSPDVEIQDQPMHDFATFENFVVVPDQ 360
 DB 301 DPESGELFALSYDVSKPYLYKFRFSPDGTGKSPDVEIQDQPMHDFATFENFVVVPDQ 360

QY 361 QVVKLPKEMIRGSGSPVYDKNKVARFGILDKYAEDSSNKWIDAPDCFCFHLNANEPEE 420
 DB 361 QVVKLPKEMIRGSGSPVYDKNKVARFGILDKYAEDSSNKWIDAPDCFCFHLNANEPEE 420

QY 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGM 480
 DB 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEPPWKVSGFAKVDLTGTVKKHLGDNRYGGEPLPLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEPPWKVSGFAKVDLTGTVKKHLGDNRYGGEPLPLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAKQV 599
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAKQV 599

RESULT 2
 Q93Z05
 ID Q93Z05 PRELIMINARY; PRT; 599 AA.
 AC Q93Z05;
 DT 01-DEC-2001 (TremBLrel. 19, Created)
 DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
 DE Putative 9-cis-epoxycarotenoid dioxygenase.
 GN MOA2.4/AT3G14440.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
 RA Kosema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene MOA2.4/AT3G14440 (GI:11994214).";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY056255; AAL07104.1; -;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 KW Dioxygenase.
 SQ SEQUENCE 599 AA; 65813 MW; B9007A2DC1C15506 CRC64;

Query Match 99.8%; Score 3143; DB 10; Length 599;
 Best Local Similarity 99.8%; Pred. No. 6.5e-237;
 Matches 599; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAAYSGRWLGNHTQPPPLSSQSSDLSYCSSLPMAASRVTRKLNYSALHTPPAL 60
 DB 1 MASFTATAAAYSGRWLGNHTQPPPLSSQSSDLSYCSSLPMAASRVTRKLNYSALHTPPAL 60

QY 61 HFPKQSSNSPAIVKPKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120
 DB 61 HFPKQSSNSPAIVKPKAKESNTKQMLFORAAAAAADAAGFLVSHEKHLPLPKTADPS 120

QY 121 VQIAGNFAFVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
 DB 121 VQIAGNFAFVNEQPVRRNLVWVKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASACRTQTNRFRVQERQLGRPVPPKAIGELHGTGIARLMLFYARAAAGIVDP 240
 DB 181 KFEHGSASACRTQTNRFRVQERQLGRPVPPKAIGELHGTGIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGRLAMSEDDLLPYQVQITPNDGLKTGVGRFDGQLESTMIAPKV 300
 DB 241 AHGTGVANAGLVYFNGRLAMSEDDLLPYQVQITPNDGLKTGVGRFDGQLESTMIAPKV 300

QY 301 DPESGELFALSYDVSKPYLYKFRFSPDGTGKSPDVEIQDQPMHDFATFENFVVVPDQ 360
 DB 301 DPESGELFALSYDVSKPYLYKFRFSPDGTGKSPDVEIQDQPMHDFATFENFVVVPDQ 360

QY 361 QVVKLPKEMIRGSGSPVYDKNKVARFGILDKYAEDSSNKWIDAPDCFCFHLNANEPEE 420
 DB 361 QVVKLPKEMIRGSGSPVYDKNKVARFGILDKYAEDSSNKWIDAPDCFCFHLNANEPEE 420

QY 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGM 480
 DB 421 TDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEPPWKVSGFAKVDLTGTVKKHLGDNRYGGEPLPLPGE 540
 DB 481 VNRNMLGRKTKFAYLALAEPPWKVSGFAKVDLTGTVKKHLGDNRYGGEPLPLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAKQV 599
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRVPGFHGTFTGADDLAKQV 599

RESULT 3
 Q9M3Z9
 ID Q9M3Z9 PRELIMINARY; PRT; 604 AA.
 AC Q9M3Z9;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE Putative 9-cis-epoxycarotenoid dioxygenase.
 GN NCED1.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Burbidge A., Taylor I.B., Thompson A.;
 RT "Putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276244; CAB76920.1; -;
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.

```

KW DIOXYGENASE.
SQ SEQUENCE 604 AA; 67288 MW; 36856BC82E1604A1 CRC64;

Query Match
Best Local Similarity 72.7%; Score 2289; DB 10; Length 604;
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;

QY 5 TATAAAGSRWLGNGHNTQPPPLSSOSSDLSYCS---SLPMASRVTRKLNVSALHTPPALH 61
DQ 3 TTTSHATNTWI-----KPKLSMPSSKEFGFASISLKNQHNQSLNINSLSQAPPLH 57
QY 62 FPKOSSN---SPAIVVVKPAKESNT---KQNLQFORAAALDAEAGFLVSHEKHLHP 113
DQ 58 FPKOSSNYQTPKNTISHPKQENNNSSSSSKWNLVQKAAAMALDAVEGALTKEHEHP 117
QY 114 PKTADPSVQIAGNAPVNEQPVRRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFD 173
DQ 118 PKTADPSVQISGNFAPVNPENPVQCSLPTGKIPKCVGVYVRNGANPLFEPTAGRHFFD 177
QY 174 DGMVHAVKFEHGSASYACRTQTNRVQBOELGRPVFPKAIHELHGHSTIARLMLFYAR 233
DQ 178 DGMVHAVQFKNGSASYACRTETETFRVQEKALGRPVFPKAIHELHGHSTIARLMLFYAR 237
QY 234 AAGIVDPAHGTGVANAGLVYFNGRLAMSDDLPLVQVQITPNDGLKTVGRFDFDGOLES 293
DQ 238 LFGLDHSRGTGVANAGLVYFNNRLAMSDDLPLVHVKVPTGDLKTEGRFDFDGOLES 297
QY 294 MIAHPKVDPSGELFALSVDVSKPYLKYPFRSPDGTKSPDVEITQDPTMMHDFATE 353
DQ 298 MIAHPKLDPSGELFALSVDVQKPYLKYPFRSKNGEKSNDVLEPVEDPTMMHDFATE 357
QY 354 FVVVPDQVQVFKLPEMIRGSPVVDKKNVAREGILDKYAEEDSNKWKIDAPDCFCFHL 413
DQ 358 FVLIPOQVQVFKSEMIRGSPVVDKKNVAREGILDKYAEEDSNKWKIDAPDCFCFHL 417
QY 414 NAWEEPETDEVVIGSCMTPDPSIFNESDENLKSLSVLEIRNLKTGSTREPIISNEDQ 473
DQ 418 NAWEEPETDEVVIGSCMTPDPSIFNECEDEGLASVLEIRNLKTGSTREPIISNEDQ 477
QY 474 VNLKAGMVRNMLGRKTKAYLALAEWPVKVSGFAKVDLTTEGVKKHLYGDNRYGGEPL 533
DQ 478 VNLKAGMVRNMLGRKTKAYLALAEWPVKVSGFAKVDLTTEGVKKHLYGDNRYGGEPL 537
QY 534 LPGE--GGEDEGYILCFVHDEKTKWSELOIVNVSLEVEATVKLPSRVPYGFHGTFFG 591
DQ 538 LPDPNSKEEDGYILAEVHDEKEWSELOIVNAMTLKLEATVKLPSRVPYGFHGTFFG 597
QY 592 DLIAKQ 597
DQ 598 NDLANQ 603

RESULT 4
O24023 PRELIMINARY; PRT; 605 AA.
ID O24023
AC O24023;
DC 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Nine-cis-epoxycarotenoid dioxygenase.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.

Query Match
Best Local Similarity 72.4%; Score 2280.5; DB 10; Length 605;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

QY 5 TATAAAGSRWLGNGHNTQPPPLSSOSSDLSYCS---SLPMASRVTRKLNVSALHTPPALH 61
DQ 3 TTTSHATNTWI-----KPKLSMPSSKEFGFASISLKNQHNQSLNINSLSQAPPLH 57
QY 62 FPKOSSN---SPAIVVVKPAKESN---TKQNLQFORAAALDAEAGFLVSHEKHLHP 112
DQ 58 FPKOSSNYQTPKNTISHPKQENNNSSSSSKWNLVQKAAAMALDAVESALTKEHEHP 117
QY 113 LPKTDPSVQIAGNAPVNEQPVRRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFD 172
DQ 118 LPKTDPSVQISGNFAPVNPENPVQCSLPTGKIPKCVGVYVRNGANPLFEPTAGRHFFD 177
QY 173 DGMVHAVKFEHGSASYACRTQTNRVQBOELGRPVFPKAIHELHGHSTIARLMLFYAR 232
DQ 178 DGMVHAVQFKNGSASYACRTETETFRVQEKALGRPVFPKAIHELHGHSTIARLMLFYAR 237
QY 233 AAGIVDPAHGTGVANAGLVYFNGRLAMSDDLPLVQVQITPNDGLKTVGRFDFDGOLES 292
DQ 238 GLFGLVDHSGTGVANAGLVYFNNRLAMSDDLPLVHVKVPTGDLKTEGRFDFDGOLES 297
QY 293 TMIAHPKVDPSGELFALSVDVSKPYLKYPFRSPDGTKSPDVEITQDPTMMHDFATE 352
DQ 298 TMIAHPKLDPSGELFALSVDVQKPYLKYPFRSKNGEKSNDVLEPVEDPTMMHDFATE 357
QY 353 NFVVPDQVQVFKLPEMIRGSPVVDKKNVAREGILDKYAEEDSNKWKIDAPDCFCFHL 412
DQ 358 NFVVPDQVQVFKSEMIRGSPVVDKKNVAREGILDKYAEEDSNKWKIDAPDCFCFHL 417
QY 413 NAWEEPETDEVVIGSCMTPDPSIFNESDENLKSLSVLEIRNLKTGSTREPIISNEDQ 472
DQ 418 NAWEEPETDEVVIGSCMTPDPSIFNECEDEGLASVLEIRNLKTGSTREPIISNEDQ 477
QY 473 QVNLKAGMVRNMLGRKTKAYLALAEWPVKVSGFAKVDLTTEGVKKHLYGDNRYGGEPL 532
DQ 478 QVNLKAGMVRNMLGRKTKAYLALAEWPVKVSGFAKVDLTTEGVKKHLYGDNRYGGEPL 537
QY 533 FLPG--GGEDEGYILCFVHDEKTKWSELOIVNVSLEVEATVKLPSRVPYGFHGTFFG 590
DQ 538 FLPRDPSKEEDGYILAEVHDEKEWSELOIVNAMTLKLEATVKLPSRVPYGFHGTFFG 597
QY 591 ADDLAQ 597
DQ 598 ANDLANQ 604

RESULT 5
O9W9F5 PRELIMINARY; PRT; 657 AA.
ID O9W9F5
AC O9W9F5;
DC 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F39.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euasterids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

Db	578	IYGEKGYGGPELFLPSGDGEDGYIMVFVHDEEKYKSELQLINAYNMKLEATVTLPSRV	633
QY	581	PYGFHGTFIGADDLAKOV 599	
		: :	
Db	638	PYGFHGTFIGKEDLSKOAL 656	
RESULT 6			
Q9FS24		PRELIMINARY;	
ID	Q9FS24	PRT:	612 AA.
AC	Q9FS24;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Neoxanthin cleavage enzyme.		
GN	CPRD65.		
OS	Vigna unguiculata (Cowpea).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.		
ON	NCBI_TaxID=4917;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20317197; Pubmed=10859185;		
RA	Iuchi S., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K.;		
RT	"A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase		
RT	involved in abscisic acid biosynthesis under water stress in drought-		
RT	tolerant cowpea."		
RL	Plant Physiol. 123:553-562(2000).		
DR	EMBL; AB030293; BABI1932.1; -.		
DR	InterPro; IPR004294; RPE65.		
DR	Pfam; PF03055; RPE65; 1.		
SO	SEQUENCE 612 AA; 67715 MW; E326A13293B188B CRC64;		

[illegible]

```

QY 521 LYCDNRYGSEPLPLPGEGGDEBGLYLCFVHDEKTKWSELOIVNAVSLVEATVKLPSPRV 580
DB 536 MYGEERKGEPLFLP-NGOKEDDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSPRV 594
QY 581 PYGFHGTFTIGADDLAKQ 597
DB 595 PYGFHGTFTHSKDLRKQ 611

RESULT 7
ID Q9M6E8 PRELIMINARY; PRT; 615 AA.
AC Q9M6E8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED1.
OS Phaseolus vulgaris (kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, TOP CROP;
RX MEDLINE=20079657; PubMed=10611388;
RA Qln X., Zeevaert J.A.;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
RT step of abscisic acid biosynthesis in water-stressed bean.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:15354-15361(1999).
DR EMBL: AF190462; AAF26356.1;
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.
KW Dioxxygenase.
SQ SEQUENCE 615 AA; 68075 MW; 0CC10F862D7DE130 CRC64;

Query Match 68.7%; Score 2165; DB 10; Length 615;
Best Local Similarity 69.2%; Pred. No. 1.7e-160;
Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

QY 11 SGRWLGNGHTQPPPLSSQSDLSYCSS-----LPMASRVTRKLN-VSSALHIPPALHPK 64
DB 6 SNTWI--NTTLPSSCSPPFDLASTSSPTTLFPKKRSSNTNTITCSLQI---LHPK 60
QY 65 Q-----SSNSPALVWPKAKE-----SNTKQ-----MNLFORAAALDAAE 101
DB 61 QYQPTSTSTTTPTPIKPTTTTTTTHRETCKPLSDTKQPPFPKWNFLQKAAATGLDVE 120
QY 102 GFLVSHKHLPLKPTADPSVOIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPL 161
DB 121 TALVSHKSHPLKPTADPKVOIAGNFAPVPEHAADQALPVVGKPKCIDGYVVRNGANPL 180
QY 162 HEPVYTHGHHFDGGMHVAHPEHGSASACRFTQTNRFVQROGLRPVFPKAIKELHGT 221
DB 181 YEPVAGHHFFDGMHVAHPEHGSASACRFTQTNRFVQROGLRPVFPKAIKELHGS 240
QY 222 GIARMLFYARAAGIIVDPAGCTGVANAGIYFNGRLLAMSEDDLPIYQVQITPNDGLKTV 281
DB 241 GIARLLFYARSLFQLVDGSHGGMVANAGIYFNNHLLAMSEDDLPIYHVRTISNGDGLTV 300
QY 282 GRFDGQLESTMTLAHPKVDPSGFLFALSVDVYVSKPKLYKFRSPDCTKSPDVEIQLDQ 341
DB 301 GRFDNGQLNSTMTLAHPKLDVNGDLHALSYDVQVKPKLYKFRSADGVKSPDVEIPLKE 360
QY 342 PTMHDFAITENFVVVPPQVVFVKLPENIRGSGPVVYDKNKVARFGILDKYAEDSSNIKW 401
DB 361 PTMHDFAITENFVVVPPQVVFVKLTMTITGSGPVVYDKNKTSRFGILDKNKAKANAEW 420
QY 402 IDAPDCCFHLWNAWEPEETDEVVIGSCMTTPDSIFNESDENLKSLSVSEIRLNKLTGES: 461
DB 421 IDAPCCFCHLWNAWEPEETDEIVVIGSCMTPADSIFNECESLSKLSVSEIRLNKLTGKS 480

```

```

QY 462 TRPPIISNEDQVNLDRAGVYNNRNLGRKTKFAYLALAEPPKYSGGFAKVDLTITGEVKKHL 521
DB 481 TRPPIIS-DAEQVNLDRAGVYNNRNLGRKTKFAYLALAEPPKYSGGFAKVDLFSGEVOKYM 539
QY 522 YGDNRYGSEPLPLPGEGGDEBGLYLCFVHDEKTKWSELOIVNAVSLVEATVKLPSPRV 581
DB 540 YGEKFGGEPLFLP-NGEEEGDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSPRV 598
QY 582 YGFHGTFTIGADDLAKQ 597
DB 599 YGFHGTFTHSKDLRKQ 614

RESULT 8
ID Q9AX24 PRELIMINARY; PRT; 625 AA.
AC Q9AX24;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED3.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, LULA;
RA Chernys J., Zeevaert J.A.D.;
RT "Abscisic acid in avocado fruit.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF224671; AAK00623.1;
DR InterPro: IPR000221; Protamine_pl.
DR Pfam: PF03055; RPE65; 1.
DR PROSITE: PS00048; PROSITE_1.
KW Dioxxygenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F33042D5F2CB6 CRC64;

Query Match 65.6%; Score 2066.5; DB 10; Length 625;
Best Local Similarity 64.9%; Pred. No. 8.9e-153;
Matches 392; Conservative 74; Mismatches 93; Indels 45; Gaps 6;

QY 38 LPMASRVTRKLNVSALHTPP-----ALHFP-----KQSS 67
DB 22 LPISKNLRSRPNFTMLKHNTPLIOCCSHSPSSSSAAVHLHLPKQPTKSKPSIKKGSST 81
QY 68 NSPATVVKPKAKESNTKQ-----MNLFORAAALDAAEGLVSH--EKLHPLPKT 116
DB 82 LTPSTIEKNPGSHOVKTQSGPNRVGPNWNIQRTAAAFALDAIEKLIARVLERHPLPKT 141
QY 117 ADPSVQIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPLHEPVTGHHFFDGM 176
DB 142 ADPEVQIAGNFAPVNEQPVRRNLVVYVKLPDSIKGYVVRNGANPLHEPVTGHHFFDGM 201
QY 177 VHAYKFEHGSASACRFTQTNRFVQROGLRPVFPKAIKELHGTGIARMLFYARAAG 236
DB 202 IHAVKFRNGSASACRFTQTNRFVQROGLRPVFPKAIKELHGTGIARMLFYARAAG 261
QY 237 IVDPAHGTGVANAGIYFNGRLLAMSEDDLPIYQVQITPNDGLKTVGRFDGQLESTMTIA 296
DB 262 LVNADEGMGVANAGIYFNGRLLAMSEDDLPIYHVRTIPSGDLKTVGRHDFDQLRSSMIA 321
QY 297 HPKVDPSGELFALSVDVYVSKPKLYKFRSPDCTKSPDVEIQLDQPTMMHDFAITENFV 356
DB 322 HPKLDPSGELFALSVDVYVSKPKLYKFRSPDCTKSPDVEIQLDQPTMMHDFAITENFV 381
QY 357 VPDQGVTFKLEPMTIRGSGPVVYDKNKVARFGILDKYAEDSSNIKWIDAPDCCFCHLWNAW 416
DB 382 IPDQGVTFKLEPMTIRGSGPVVYDKNKVARFGILDKYAEDSSNIKWIDAPDCCFCHLWNAW 441

```

```

QY 417 EBPETDEVVIGSCMTPPDSIFNESDENLKSVLSEIRLNKLTGSTRRIISNEDQVNL 476
Db 442 EPESEGVVVGSCMTPPDSIFNENESLSKLTIRLNTRIGESTRRRII-DPQRLNL 500
QY 477 EAGMVRNRLGKRTKTFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPLFLPG 536
Db 501 EAGMVRNRLGKRTKTFAYLALAEPPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPLFLPG 560
QY 537 E---GGEDEGYILCFVHDEKWKSELOLVNVAISLEVEATVKLPSRVPYGFHGTFTGADD 593
Db 561 EYSTSGREDDGVVGFHDEKTSRSELLINANNRLEASVMLPSRVPYGFHGTFTISSKD 620
QY 594 LAQK 597
Db 621 LAQK 624

RESULT 9
Q9C621
AC Q9C621; PRELIMINARY; PRT; 589 AA.
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase, putative.
GN T2H7.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li X.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT *Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.
RL Nature 408:816-820(2000).
DR EMBL; AC074176; AAG50855.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 589 AA; 65336 MW; 6A79B2BF7EAA8179 CRC64;

Query Match 64.9%; Score 2044.5; DB 10; Length 589;
Best Local Similarity 68.1%; Pred. No. 4.2e-151;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

QY 27 SSSLSVSCSLPMSRVTRKLNLSALHTPPALHFFKSSNSPAIVVKKPKAKESNTKQM 86
Db 27 SPSSSVSEFTNKPR----RRKLSANSVDTNLLNPNVPSNPPII-----PEKDTSRW 76
QY 87 NLFQRAAALDAAGFLYSHEKHLPLKPTADPSVQIAGNFAPVNEQPVRRNLVVGKLP 146
Db 77 NPLQRAAALDAETALLRRERSKRLPKTVDPRHQISGNVAPVPEQSVKSSLSVDGKIP 136
QY 147 DSIKGVYVRNGANPLPEVPTGTHHFFDGDGMVHAVKFEHGSASACRFTTQTNFVQERQLG 206

```

```

Db 137 DCTDGYLVRNGANPLPEVPTGTHHFFDGDGMVHAVKTINGDASYSCTFETRLVQEKQLG 196
QY 207 RVFPAKAIAGELGHTGIARIMLFYARAAAGIVDPAHGTGVANAGLVYFNGRIIAMSDDL 266
Db 197 SPIFPAKAIAGELGHTGHTGIARIMLFYARGLFGLLNKNGTGVANAGLVYFHDRLAMSDDL 256
QY 267 PYOVLTTPNGDLTKVGRFDFDQLESTMTAHKVDPESELGALSVDVYVKPKYLFYFES 326
Db 257 PYOVRTDNGDLTKVGRFDFDQLESTMTAHKVDPESELGALSVDVYVKPKYLFYFES 316
QY 327 PDGTSKPDVEIQLODPTMMHDAITENFVVPDQVQVFKLPEMIRGSGPVVYDKNKVARF 386
Db 317 PEGEKSPDVEIQLODPTMMHDAITENFVVPDQVQVFKLPEMIRGSGPVVYDKNKVARF 376
QY 387 GILDYAEUSSNKKWIDAPDCFCFHLWNABEPETDEVVIGSCMTPPDSIFNESDENLJ 446
Db 377 GILPRNAKDAEMVWVESPTFCFHLWNABEPETDEVVIGSCMTPPDSIFNESDENLJ 436
QY 447 SVLSEIRLNKLTGSTRRIISNEDQVNLKAGMVRNRLGKRTKTFAYLALAEPPKVS 506
Db 437 SVLSEIRLNKLTGSTRRIISNEDQVNLKAGMVRNRLGKRTKTFAYLALAEPPKVS 495
QY 507 FAKVDLTGTEVKKHLYGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKWKSELOIYN 564
Db 496 FAKVDLTGTEVKKHLYGDNRYGGEPLFLPG--EGGEDEGYILCFVHDEKWKSELOIYN 555
QY 565 AVSLEVEATVKLPSRVPYGFHGTFTGADDLAQK 597
Db 556 AVTLEATVKLPSRVPYGFHGTFTVNSADMLNQ 588

RESULT 10
Q49505
ID Q49505; PRELIMINARY; PRT; 583 AA.
AC Q49505;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE NEOXANTHIN cleavage enzyme-like protein.
GN F28J12.10 OR AT4G18350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021710; CAAL16715.1;
DR EMBL; AL161548; CAB78837.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
SQ SEQUENCE 583 AA; 65066 MW; A138F93542E50852 CRC64;

Query Match 63.2%; Score 1991; DB 10; Length 583;
Best Local Similarity 64.5%; Pred. No. 6.3e-147;
Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

```



```

QY 75 -RFAKESNTKOMNLFQRAAAALDA-ABGFLVS-BEKLHPLKTDPSVQIAGNEAPVN 131
Db 76 AARPKAEGGKKQLNLFQRAAAALDA-ABGFLVS-BEKLHPLKTDPSVQIAGNEAPVN 131
QY 132 EQPVRENLPWVKGLPSIKGVYVRNGANLPHVPTGHHFFDGDGMVHAKVFEHGA-SYA 190
Db 133 ERFVHELPRVSGRIPFIDGVYARNGANCFDPVAGHHLFDGDMVHALKIRNGAAESYA 191
QY 191 CRETOQNRFRQVQOLGRPVFPKPAIGELHGHGTGIALMLFYARAAAGTVPDAGHTGTVANAG 250
Db 192 CRETEARLQERARIGRPFPKPAIGELHGHGTGIALMLFYARAAAGTVPDAGHTGTVANAG 251
QY 251 LVYFNGRLAMSEDDLPYQVQITPNDGLKTVGRFDDGDLSTMIABPKVDPESGELFAL 310
Db 252 LVYFNGRLAMSEDDLPYHVRVADDDGLETVGRYDFDGLGCGAMIAHPKLPDPAIGELHAL 311
QY 311 SYDVVSKPKYKFRFPDGTGKSPDVLQIDOPTMMHDFAITENFVVPDQOVVFKLPPEMI 370
Db 312 SYDVIRPKYKFRFPDGTGKSDVEIPLEQPTMIHDFAITENFVVPDQOVVFKLQEML 371
QY 371 RGGSPVYVYDKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPEETDEVVYIGSC 430
Db 372 RGGSPVYVYDKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPEETDEVVYIGSC 431
QY 431 MPPDIFNESDENLKSIVSEIRLNKLTGSTRPPIISNEDQOVNLEAGMVRNMLGRKT 490
Db 432 MPPDIFNESDENLKSIVSEIRLNKLTGSTRPPIISNEDQOVNLEAGMVRNMLGRKT 490
QY 491 KPAYLALAPKPKVSGFAKVDLTGTVKHKLYGDNRYGGEPLFLPGEGE-DEEGY 545
Db 491 RYAYLAVAPKPKVSGFAKVDLTGTVKHKLYGDNRYGGEPLFLPGEGE-DEEGY 545
QY 546 ILCFVHDEKTSKSELQIVNAVSLVEATVKLPSRVPYGFHGTFIGADDLAKQ 597
Db 551 VLTVEHDERAGRSLLVWNAADIRLEATVQLPSRVDFGFGHTFIGQELEAQ 602

RESULT 13
Q9LRM7 ID Q9LRM7 PRELIMINARY; PRT: 577 AA.
AC Q9LRM7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 9-cls-epoxycarotenoid (Mousserear).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB028621; BAB01363.1;
DR InterPro; IPR004294; RPF65.
DR Pfam; PF03055; RPF65; 1.
KW Dioxigenase.
SQ SEQUENCE 577 AA; 63821 MW; 515E0A1B321B1FE1 CRC64;

```

Query Match 52.8%; Score 1663.5; DB 10; Length 577;
 Best Local Similarity 55.4%; Pred. No. 2.4e-121;
 Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

```

QY 24 LSSQSSDLSYCSLPMASVTRKLVSS-ALHTFPALHPKQSSNSPATVVKPKAKESN 82
Db 9 LLPTKTSRSHLLPQPKNANISRIILNPKIPTLPDLTSPVSP-----VKLKPTPN 62
QY 83 TKOMNLFQRAAAALDAAB-GEIVSHKELHPLKTDPSVQIAGNEAPVNEQVRENLPV 141
Db 63 ---LNLQKLAATMLDKIESSIVPMQONRPLKPDPAVLQSGNFAVNECEPVQNGLEV 119
QY 142 VGKLPDSIKGVYVRNGANLPHVPTGHHFFDGDGMVHAKV--PEHGSASVACRFTOTNR 199
Db 120 VGQIPSLKGVYVRNGANLPHVPTGHHFFDGDGMVHAKV--PEHGSASVACRFTOTNR 178
QY 200 VQERQLGRPVFPKPAIGELHGHGTGIALMLFYARAAAGTVPDAGHTGTVANAGLVYFNGRL 259
Db 179 VOETALGRSVFPKPAIGELHGHGTGIALMLFYARAAAGTVPDAGHTGTVANAGLVYFNGRL 238
QY 260 AMSEDDLPYQVQITPNDGLKTVGRFDDGDLSTMIABPKVDPESGELFALSVDVYVSKPY 319
Db 239 AMSEDDLPYQVQITPNDGLKTVGRFDDGDLSTMIABPKVDPESGELFALSVDVYVSKPY 298
QY 320 LKTFRSPDGTGKSPDVLQIDOPTMMHDFAITENFVVPDQOVVFKLPPEMI 379
Db 299 LRYLKENTCGKTRDVEITLPEPTMIHDFAITENFVVPDQOVVFKLPPEMI 358
QY 380 KKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPEETDEVVYIGSC 435
Db 359 KKNVAREGIIDKYAEDSSNKKWIDAPDCFCFHLNNAWEEPEETDEVVYIGSC 417
QY 436 SIPNESDENLKSIVSEIRLNKLTGSTRPPIISNEDQOVNLEAGMVRNMLGRKTFAVL 495
Db 418 TIFSEGEPTVELSEIRLNKLTGSTRPPIISNEDQOVNLEAGMVRNMLGRKTFAVL 473
QY 496 ALAEPKPKVSGFAKVDLTGTVKHKLYGDNRYGGEPLFLPGEGE-DEEGY 545
Db 474 ALAEPKPKVSGFAKVDLTGTVKHKLYGDNRYGGEPLFLPGEGE-DEEGY 533
QY 556 WKSELQIVNAVSLVEATVKLPSRVPYGFHGTFIGADDLAKQ 598
Db 534 DESEFVVVDATDMKQVAAVRLPERVYGFHGTFIGADDLAKQ 576

RESULT 14
O49675 ID O49675 PRELIMINARY; PRT: 595 AA.
AC O49675;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NEOXANTHIN cleavage enzyme-like protein (AT4G19170/T18B16_140).
GN T18B16.140 OR AT4G19170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB028621; BAB01363.1;
DR InterPro; IPR004294; RPF65.
DR Pfam; PF03055; RPF65; 1.
KW Dioxigenase.
SQ SEQUENCE 577 AA; 63821 MW; 515E0A1B321B1FE1 CRC64;

```

